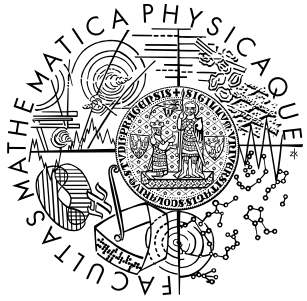
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Dita Rensová

Klasifikační analýza

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Dr.rer.nat. Jan Kalina
Studijní program: Matematika
Studijní obor: Obecná matematika

2008

Ráda bych poděkovala vedoucímu mé bakalářské práce Dr.rer.nat. Janu Kalinovi za cenné rady a ochotu, se kterou mi věnoval svůj čas, a mému otci Ing. Pavlu Rensovi za poskytnutí dat a informace potřebné k jejich zpracování.

Prohlašuji, že jsem svou bakalářskou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 6.8.2008 Dita Rensová

Obsah

Název práce: Klasifikační analýza
Autor: Dita Rensová
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: Dr.rer.nat. Jan Kalina
e-mail vedoucího: kalina@karlin.mff.cuni.cz

Abstrakt: V této práci se zabýváme modely klasifikační analýzy. Popíšeme jednotlivá klasifikační pravidla a souvislosti mezi nimi. Nejprve se zaměříme na modely lineární a kvadratické klasifikace pro případ dvou skupin, které dále zobecníme na lineární a kvadratické modely pro případ klasifikace do více skupin. Poté se budeme zabývat pravděpodobností špatné klasifikace určitého objektu do skupiny a metodami, jak tuto pravděpodobnost odhadnout. Dále se zmíníme o využití diskriminačních skórů při klasifikaci a seznámíme se s modelem logistické klasifikace. Na závěr předvedeme použití některých vybraných modelů na konkrétních datech z oboru lesnictví.

Klíčová slova: Lineární klasifikace, kvadratická klasifikace, logistická klasifikace, diskriminace a klasifikace

Title: Classification analysis
Author: Dita Rensová
Department: Department of Probability and Mathematical Statistics
Supervisor: Dr.rer.nat. Jan Kalina
Supervisor's e-mail address: kalina@karlin.mff.cuni.cz

Abstract: In the present work we study methods for classification analysis. We describe some classification rules and show connections between them. First, we study models of linear and quadratic classification into two groups. Then we generalize these models for classification into several groups. We study the probability of missclassification of some object and describe some methods to estimate this probability. We also apply discriminant scores to the classification problem and further describe logistic classification. Finally we illustrate the usage of the methods on a data set from forestry.

Keywords: Linear classification, quadratic classification, logistic classification, discrimination and classification

Kapitola 1

Úvod

V běžném životě se setkáváme s tím, že objekty nebo jedince nějaké populace můžeme zařadit do různých skupin. Lékaři například diagnostikují, zda pacient trpí nebo netrpí určitou chorobou, žáky ve třídě můžeme rozdělit na chlapce a dívky, nebo je můžeme rozdělit do skupin třeba podle jejich oblíbeného předmětu. Stromy v lese můžeme rozdělit podle dřevin a celé lesní porosty můžeme zase rozdělit do skupin například podle množství dřeva, které vyprodukují. Dalších příkladů by se našla jistě celá řada. V některých situacích je naše rozhodnutí, do které skupiny daného jedince zařadit, jednoduché, jindy se rozhodujeme na základě hlubšího pozorování, výsledků různých testů, měření či výpočtů.

Většinou na každém objektu ve všech skupinách pozorujeme celou řadu veličin. Objekty z jedné skupiny vykazují stejné nebo podobné výsledky, zatímco výsledky zjištěné na objektech z různých skupin by se měly lišit. Pokud se objeví nějaký nový objekt, o kterém nevíme, do jaké skupiny patří, zařadíme ho přirozeně do skupiny objektů s podobnými výsledky. Počet veličin, které na objektech pozorujeme, může být velký. Oproti tomu rozdíly ve výsledcích mohou být nepatrné, a proto není vždy jednoduché daný objekt správně zařadit. Klasifikační analýza je matematickým nástrojem pro řešení těchto problémů. V dalším textu budeme předpokládat, že veličiny, které na objektech pozorujeme, jsou ovlivněny náhodou. Uvedeme si některé metody, jak rozlišit chování těchto veličin v závislosti na tom, z jaké skupiny objekty pochází, a jak sestavit pravidla pro klasifikaci dosud nazařazených objektů.

Cílem této práce je podat přehledný popis jednotlivých klasifikačních pravidel a vysvětlit jejich vzájemné souvislosti. Ve druhé kapitole se budeme zabývat případem, kdy máme pouze dvě skupiny objektů a v kapitole třetí

rozšíříme naše poznatky na klasifikaci do více skupin. Ve čtvrté kapitole se přesvědčíme, že klasifikace založená na diskriminačních skórech dává ekvivalentní klasifikační pravidla. V šesté kapitole potom zmíníme využití logistické regrese při klasifikaci a na závěr si předvedeme použití některých popsaných metod na konkrétních datech.

Kapitola 2

Klasifikace do dvou skupin

V této a následující kapitole budeme vycházet především z knihy [6]. Pro začátek se zaměříme na případ, kdy máme pouze dvě skupiny objektů, označme je π_1 a π_2. Předpokládejme, že na každém objektu pozorujeme p-rozměrný náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_p)'$. Na základě zjištěných hodnot $\boldsymbol{X} = \boldsymbol{x}$ chceme daný objekt zařadit do skupiny π_1 nebo π_2. (Říkáme také, že pozorování $\boldsymbol{x}$ klasifikujeme jako π_1 nebo π_2.) V této kapitole, stejně jako v kapitolách 3 a 4, budeme předpokládat, že rozdělení $\boldsymbol{X}$ je v obou skupinách absolutně spojité. Označme $f_1(\boldsymbol{x})$ (resp. $f_2(\boldsymbol{x})$) hustotu rozdělení $\boldsymbol{X}$, pokud pozorování pochází z první (resp. z druhé) skupiny. Nyní rozlišíme dva případy.

2.1 Stejné varianční matice

V tomto odstavci budeme předpokládat, že varianční matice $\boldsymbol{\Sigma}_1$ a $\boldsymbol{\Sigma}_2$ první a druhé skupiny jsou stejné, označme $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. Dále označme p_1 (resp. p_2) apriorní pravděpodobnosti, že pozorování $\boldsymbol{x}$ bude pocházet z první (resp. z druhé) skupiny. Pravidlo, podle kterého zařadíme pozorování $\boldsymbol{x}$ do první nebo druhé skupiny zvolíme takové, aby bylo v jistém smyslu optimální. Protože bychom rádi zařadili co nejvíce objektů do správné skupiny, za optimální klasifikační pravidlo vybereme to, které minimalizuje pravděpodobnost špatné klasifikace.

Pravidlo 1 *Jestliže*

$$p_1 f_1(\boldsymbol{x}) > p_2 f_2(\boldsymbol{x}), \tag{2.1}$$

pak klasifikujeme $\boldsymbol{x}$ jako π_1. V opačném případě klasifikujeme $\boldsymbol{x}$ jako π_2.

Věta 1 *Nechť rozdělení ve skupinách π_i mají hustotu $f_i(\boldsymbol{x})$, $i = 1, 2$. Potom při použití klasifikačního pravidla 1 je minimalizována pravděpodobnost špatné klasifikace.*

Důkaz. Každé klasifikační pravidlo určuje rozklad $\mathbb{R}_p$ na dvě disjunktní množiny $U_1, U_2 \subset \mathbb{R}_p$, následujícím způsobem:

$$\boldsymbol{x} \in U_i \Leftrightarrow \boldsymbol{x} \text{ je klasifikováno jako } \pi_i,\ i = 1, 2.$$

Označme $P_{U_1,U_2}(\boldsymbol{x})$ pravděpodobnost toho, že $\boldsymbol{x}$ bude špatně klasifikováno při použití klasifikačního pravidla s rozkladem U_1, U_2. Podobně označme podmíněnou pravděpodobnost

$$P_{U_1,U_2}(i|j) = P_{U_1,U_2}(\boldsymbol{x} \text{ klasifikováno jako } \pi_i | \boldsymbol{x} \text{ pochází z } \pi_j), \quad i, j = 1, 2.$$

Definujme

$$W_1 = \{\boldsymbol{x} \in \mathbb{R}_p : p_1 f_1(\boldsymbol{x}) > p_2 f_2(\boldsymbol{x})\},$$
$$W_2 = \{\boldsymbol{x} \in \mathbb{R}_p : p_1 f_1(\boldsymbol{x}) \leq p_2 f_2(\boldsymbol{x})\}.$$

W_1 a W_2 tvoří disjunktní rozklad $\mathbb{R}_p$, který je určen klasifikačním pravidlem (2.1). Ukážeme, že při tomto rozkladu je minimalizována pravděpodobnost špatné klasifikace. Jinak řečeno ukážeme, že pro jiné klasifikační pravidlo, reprezentované rozkladem $X_1, X_2 \subset \mathbb{R}_p$, kde $X_1 \cup X_2 = \mathbb{R}_p$ a $X_1 \cap X_2 = \emptyset$, je $P_{X_1,X_2}(\boldsymbol{x}) \geq P_{W_1,W_2}(\boldsymbol{x})$.

$$\begin{aligned}
P_{X_1,X_2}(\boldsymbol{x}) &= p_1 P_{X_1,X_2}(2|1) + p_2 P_{X_1,X_2}(1|2) \\
&= p_1 \int_{X_2} f_1(\boldsymbol{x})\mathrm{d}x + p_2 \int_{X_1} f_2(\boldsymbol{x})\mathrm{d}x \\
&= p_1\Big(\int_{X_2\cap W_1} f_1(\boldsymbol{x})\mathrm{d}x + \int_{X_2\cap W_2} f_1(\boldsymbol{x})\mathrm{d}x\Big) + p_2\Big(\int_{X_1\cap W_1} f_2(\boldsymbol{x})\mathrm{d}x + \int_{X_1\cap W_2} f_2(\boldsymbol{x})\mathrm{d}x\Big) \\
&= \int_{X_2\cap W_1} p_1 f_1(\boldsymbol{x})\mathrm{d}x + \int_{X_2\cap W_2} p_1 f_1(\boldsymbol{x})\mathrm{d}x + \int_{X_1\cap W_1} p_2 f_2(\boldsymbol{x})\mathrm{d}x + \int_{X_1\cap W_2} p_2 f_2(\boldsymbol{x})\mathrm{d}x \\
&\geq \int_{X_2\cap W_1} p_2 f_2(\boldsymbol{x})\mathrm{d}x + \int_{X_2\cap W_2} p_1 f_1(\boldsymbol{x})\mathrm{d}x + \int_{X_1\cap W_1} p_2 f_2(\boldsymbol{x})\mathrm{d}x + \int_{X_1\cap W_2} p_1 f_1(\boldsymbol{x})\mathrm{d}x \\
&= p_1 \int_{W_2} f_1(\boldsymbol{x})\mathrm{d}x + p_2 \int_{W_1} f_2(\boldsymbol{x})\mathrm{d}x \\
&= p_1 P_{W_1,W_2}(2|1) + p_2 P_{W_1,W_2}(1|2) \\
&= P_{W_1,W_2}(\boldsymbol{x})
\end{aligned}$$

□

Nyní aplikujeme toto optimální pravidlo na mnohorozměrné normální rozdělení. (Pro definici mnohorozměrného normálního rozdělelní viz definice (A.1).)

Věta 2 *Nechť skupiny π_i mají normální rozdělení $\mathsf{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 1, 2$. Potom (2.1) je ekvivalentní s*

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} > \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \ln\frac{p_2}{p_1}. \tag{2.2}$$

Důkaz. Nerovnost (2.1) je ekvivalentní s nerovností

$$\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} > \frac{p_2}{p_1},$$

kterou dále zlogaritmujeme

$$\ln\left[\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}\right] > \ln\frac{p_2}{p_1}.$$

Do levé strany dosadíme za $f_1(\boldsymbol{x})$ a $f_2(\boldsymbol{x})$ hustoty normálního rozdělení (A.1), takže

$$\ln\left[\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}\right] = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_2). \tag{2.3}$$

Využijeme toho, že matice $\boldsymbol{\Sigma}^{-1}$ je symetrická, a tedy pro libovolná $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_p$ platí $(\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y})' = \boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}$. Protože $\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} \in \mathbb{R}$, platí navíc $(\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y})' = \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$. Rovnost (2.3) můžeme tedy dále upravit

$$\begin{aligned}
\ln\left[\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}\right] &= \frac{1}{2}(-\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 \\
&\quad + \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2) \\
&= \frac{1}{2}[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\
&\quad - \boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1] \\
&= \frac{1}{2}[2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)] \\
&= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2).
\end{aligned}$$

Dostáváme tak

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) > \ln\frac{p_2}{p_1}$$

a to už je ekvivalentní s (2.2). □
Můžeme tedy formulovat variantu klasifikačního pravidla pro případ mnohorozměrného normálního rozdělení.

Pravidlo 2 *Jestliže*

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} > \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \ln\frac{p_2}{p_1}, \tag{2.4}$$

potom klasifikujeme $\boldsymbol{x}$ *jako* π_1*, jinak klasifikujeme* $\boldsymbol{x}$ *jako* π_2*.*

Pro zajímavost se podívejme, jak vypadá pravidlo 2 pro jeden speciální případ. Předpokládejme, že normální rozdělení ve skupinách π_i, $i = 1, 2$ je jednorozměrné s parametry (μ_i,σ^2) a navíc je splněna podmínka $p_1 = p_2 = \frac{1}{2}$. V tomto případě má (2.4) tvar

$$\frac{(\mu_1 - \mu_2)}{\sigma^2}x > \frac{1}{2}\frac{(\mu_1 - \mu_2)(\mu_1 + \mu_2)}{\sigma^2},$$

odkud už dostaneme

$$x > \frac{(\mu_1 + \mu_2)}{2}, \text{ pokud } \mu_1 > \mu_2,$$

$$x < \frac{(\mu_1 + \mu_2)}{2}, \text{ pokud } \mu_1 < \mu_2.$$

V tomto konkrétním případě je tak klasifikační pravidlo jednoduché. Pozorování x zařadíme do té skupiny, k jejíž střední hodnotě leží blíž.

Hodnoty $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ a $\boldsymbol{\Sigma}$ často nejsou známy. V takovém případě se při sestavování klasifikačních pravidel využívá odhadů, získaných na základě již zařazených pozorování. Předpokládejme, že máme náhodný výběr z rozdělení π_1 o rozsahu n_1, $\boldsymbol{X}_1 = (\boldsymbol{X}_1^1, \boldsymbol{X}_1^2, \ldots, \boldsymbol{X}_1^{n_1})'$, a náhodný výběr z rozdělení π_2 o rozsahu n_2, $\boldsymbol{X}_2 = (\boldsymbol{X}_2^1, \boldsymbol{X}_2^2, \ldots, \boldsymbol{X}_2^{n_2})'$. Toto značení bude výhodné pro formulaci a vysvětlení jednotlivých klasifikačních pravidel. Zdůrazněme, že $\boldsymbol{X}_1$ je matice typu $n_1 \times p$, jejíž řádky tvoří p-rozměrná pozorování z rozdělení π_1. Podobně $\boldsymbol{X}_2$ je matice typu $n_2 \times p$, jejíž řádky tvoří p-rozměrná

pozorování z rozdělení π_2. Vypočítáme výběrové průměry a výběrové variančí matice (pozorované hodnoty nyní značíme malými písmeny)

$$\bar{\boldsymbol{x}}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} \boldsymbol{x}_1^j; \qquad \boldsymbol{S}_1 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1} (\boldsymbol{x}_1^j - \bar{\boldsymbol{x}}_1)(\boldsymbol{x}_1^j - \bar{\boldsymbol{x}}_1)'$$

$$\bar{\boldsymbol{x}}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} \boldsymbol{x}_2^j; \qquad \boldsymbol{S}_2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2} (\boldsymbol{x}_2^j - \bar{\boldsymbol{x}}_2)(\boldsymbol{x}_2^j - \bar{\boldsymbol{x}}_2)'.$$

Jako odhad $\boldsymbol{\mu}_i$ použijeme $\bar{\boldsymbol{x}}_i$, $i = 1, 2$, a protože předpokládáme, že oba výběry pochází z rozdělení se stejnou varianční maticí $\boldsymbol{\Sigma}$, použijeme jako její odhad sdruženou výběrovou varianční matici

$$\boldsymbol{S} = \frac{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2}{n_1 + n_2 - 2}.$$

Dosadíme-li tyto odhady do (2.2), získáme výběrovou analogii klasifikačního pravidla 2.

Pravidlo 3 *Jestliže*

$$(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}\boldsymbol{x} > \frac{1}{2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) + \ln\frac{p_2}{p_1}, \tag{2.5}$$

pak pozorování $\boldsymbol{x}$ klasifikujeme jako π_1, jinak klasifikujeme $\boldsymbol{x}$ jako π_2.

Protože levá strana nerovnice (2.5) je lineární funkcí složek $\boldsymbol{x}$, nazývá se pravidlo 3 *lineární klasifikační pravidlo* a funkce $L(\boldsymbol{x}) = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}\boldsymbol{x}$ se nazývá *lineární klasifikační funkce*.

Podívejme se nyní na to, jak je to s optimalitou lineárního klasifikačního pravidla 3. Tentokrát označíme pozorované hodnoty opět velkými písmeny, abychom zdůraznili, že se jedná o náhodné veličiny. Při značení $\bar{\boldsymbol{X}}_1 = (\bar{X}_{11}, \bar{X}_{12}, \ldots, \bar{X}_{1p})'$ a $\boldsymbol{\mu}_1 = (\mu_{11}, \mu_{12}, \ldots, \mu_{1p})$, dostáváme z důsledku A.5, že pro $n \to \infty$ konverguje $\bar{X}_{1j}$ v pravděpodobnosti k μ_{1j} pro každé $j = 1, 2, \ldots, p$, a tedy i $\bar{\boldsymbol{X}}_1$ konverguje v pravděpodobnosti k $\boldsymbol{\mu}_1$. Stejně tak $\bar{\boldsymbol{X}}_2$ konverguje v pravděpodobnosti k $\boldsymbol{\mu}_2$. Ukažme si, že i $\boldsymbol{S}_i$ konvergují v pravděpodobnosti k $\boldsymbol{\Sigma}$. Při značení $\boldsymbol{\Sigma} = (\sigma_{jk})$ a $\boldsymbol{S}_i = (s_{jk}^i)$, $j, k =$

$1, 2, \ldots, p$, $i = 1, 2$, platí

$$\begin{aligned} s^1_{jk} &= \frac{1}{n_1 - 1} \sum_{l=1}^{n_1} (X^l_{1j} - \bar{X}_{1j})(X^l_{1k} - \bar{X}_{1k}) \\ &= \frac{1}{n_1 - 1} \sum_{l=1}^{n_1} (X^l_{1j} - \mu_{1j} + \mu_{1j} - \bar{X}_{1j})(X_{1k}l - \mu_{1k} + \mu_{1k} - \bar{X}_{1k}) \\ &= \frac{1}{n_1 - 1} \sum_{l=1}^{n_1} (X^l_{1j} - \mu_{1j})(X^l_{1k} - \mu_{1k}) + \frac{n_1}{n_1 - 1}(\bar{X}_{1j} - \mu_{1j})(\bar{X}_{1k} - \mu_{1k}). \end{aligned}$$

Z důsledku A.5 vidíme, že druhý člen v tomto vyjádření s^1_{jk} konverguje v pravděpodobnosti k nule a první člen konverguje v pravděpodobnosti k $\mathsf{E}(\bar{X}_{1j} - \mu_{1j})(\bar{X}_{1k} - \mu_{1k}) = \sigma_{jk}$. Matice $\boldsymbol{S}_1$ tedy konverguje k varianční matici $\boldsymbol{\Sigma}$, podobně i matice $\boldsymbol{S}_2$. Jinak řečeno, pro dostatečně velké rozsahy výběrů n_i se odhady $\bar{\boldsymbol{X}}_i$ a $\boldsymbol{S}_i$ málo liší od skutečných hodnot $\boldsymbol{\mu}_i$ a $\boldsymbol{\Sigma}$ a lineární klasifikační pravidlo je asymptoticky optimální.

Často také neznáme hodnoty apriorních pravděpodobností p_1 a p_2. Nabízí se možnost odhadnout p_i relativními četnostmi objektů ze skupiny π_i ve sdruženém výběru $(\boldsymbol{X}'_1, \boldsymbol{X}'_2)'$, tj. za p_i dosadíme v jednotlivých klasifikačních pravidlech hodnoty $\frac{n_i}{n_1+n_2}$, $i = 1, 2$. Tento postup se však doporučuje pouze v případě, pokud víme, že zastoupení objektů z obou skupin v celém výběru odpovídá reálné situaci. Pokud tomu tak není, volí se $p_1 = p_2 = \frac{1}{2}$. Zabývejme se nyní tímto případem, kdy $p_1 = p_2$.

Věta 3 *Jestliže $p_1 = p_2$, potom nerovnost (2.5) je ekvivalentní s nerovností*

$$\boldsymbol{D}_1^2 < \boldsymbol{D}_2^2, \text{ kde } \boldsymbol{D}_i^2 = (\boldsymbol{x} - \bar{\boldsymbol{x}}_i)'\boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_i),\ i = 1, 2.$$

Důkaz.

$$\begin{aligned} \boldsymbol{D}_1^2 &< \boldsymbol{D}_2^2 \\ (\boldsymbol{x} - \bar{\boldsymbol{x}}_1)'\boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1) &< (\boldsymbol{x} - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_2) \\ \boldsymbol{x}'\boldsymbol{S}^{-1}\boldsymbol{x} - \bar{\boldsymbol{x}}_1'\boldsymbol{S}^{-1}\boldsymbol{x} - \boldsymbol{x}'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_1'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}_1 &< \boldsymbol{x}'\boldsymbol{S}^{-1}\boldsymbol{x} - \bar{\boldsymbol{x}}_2'\boldsymbol{S}^{-1}\boldsymbol{x} \\ &\quad -\boldsymbol{x}'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}_2 + \bar{\boldsymbol{x}}_2'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}_2 \\ \bar{\boldsymbol{x}}_1'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}_2 - \bar{\boldsymbol{x}}_2'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}_1 &< 2\bar{\boldsymbol{x}}_1'\boldsymbol{S}^{-1}\boldsymbol{x} - 2\bar{\boldsymbol{x}}_2'\boldsymbol{S}^{-1}\boldsymbol{x} \\ \frac{1}{2}[\bar{\boldsymbol{x}}_1'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) - \bar{\boldsymbol{x}}_2'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_2 + \bar{\boldsymbol{x}}_1)] &< (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}\boldsymbol{x} \\ \frac{1}{2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) &< (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}\boldsymbol{x} \end{aligned}$$

Protože pro $p_1 = p_2$ je $\ln \frac{p_2}{p_1} = 0$, je poslední nerovnost ekvivalentní s (2.5). □

Klasifikační pravidlo pro $p_1 = p_2$ má tedy následující tvar.

Pravidlo 4 *Jestliže*

$$\boldsymbol{D}_1^2 < \boldsymbol{D}_2^2,$$

pak klasifikujeme $\boldsymbol{x}$ *jako* π_1*, jinak klasifikujeme* $\boldsymbol{x}$ *jako* π_2*.*

2.2 Různé varianční matice

Nyní budeme uvažovat situaci, kdy varianční matice rozdělení π_1 a π_2 jsou různé, $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$. Podívejme se, jak v takovém případě vypadá klasifikační pravidlo 1 pro mnohorozměrné normální rozdělení. Stejně jako v důkazu věty 2 přejdeme od (2.1) k ekvivalentní nerovnosti

$$\ln \Big[\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}\Big] > \ln \frac{p_2}{p_1}.$$

Levou stranu nerovnice dále upravíme

$$\begin{aligned}
Q(\boldsymbol{x}) &= \ln \Big[\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}\Big] \\
&= \ln \Big\{ \frac{\exp\big\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\big\}/[(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma_1}|^{\frac{1}{2}}]}{\exp\big\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)\big\}/[(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma_2}|^{\frac{1}{2}}]} \Big\} \\
&= \frac{1}{2}\ln\Big(\frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|}\Big) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)'{\boldsymbol{\Sigma}_1}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) \\
&\quad + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)'{\boldsymbol{\Sigma}_2}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2) \\
&= \frac{1}{2}\ln\Big(\frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|}\Big) - \frac{1}{2}\big(\boldsymbol{x}'{\boldsymbol{\Sigma}_1}^{-1}\boldsymbol{x} - \boldsymbol{\mu}_1'{\boldsymbol{\Sigma}_1}^{-1}\boldsymbol{x} - \boldsymbol{x}'{\boldsymbol{\Sigma}_1}^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_1'{\boldsymbol{\Sigma}_1}^{-1}\boldsymbol{\mu}_1 \\
&\quad - \boldsymbol{x}'{\boldsymbol{\Sigma}_2}^{-1}\boldsymbol{x} + \boldsymbol{\mu}_2'{\boldsymbol{\Sigma}_2}^{-1}\boldsymbol{x} + \boldsymbol{x}'{\boldsymbol{\Sigma}_2}^{-1}\boldsymbol{\mu}_2 - \boldsymbol{\mu}_2'{\boldsymbol{\Sigma}_2}^{-1}\boldsymbol{\mu}_2\big) \\
&= \frac{1}{2}\ln\Big(\frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|}\Big) - \frac{1}{2}(\boldsymbol{\mu}_1'{\boldsymbol{\Sigma}_1}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2'{\boldsymbol{\Sigma}_2}^{-1}\boldsymbol{\mu}_2) \\
&\quad + (\boldsymbol{\mu}_1'{\boldsymbol{\Sigma}_1}^{-1} - \boldsymbol{\mu}_2'{\boldsymbol{\Sigma}_2}^{-1})\boldsymbol{x} - \frac{1}{2}\boldsymbol{x}'({\boldsymbol{\Sigma}_1}^{-1} - {\boldsymbol{\Sigma}_2}^{-1})\boldsymbol{x}.
\end{aligned}$$

Z poslední rovnosti vidíme, že $Q(\boldsymbol{x})$ obsahuje druhé mocniny a součiny složek vektoru $\boldsymbol{x}$, a proto se tato funkce nazývá *kvadratická klasifikační funkce.*

Z pravidla 1 a toho, co jsme si právě ukázali, můžeme formulovat klasifikační pravidlo pro případ různých variančních matic.

Pravidlo 5 *Jestliže*

$$Q(\boldsymbol{x}) > \ln \frac{p_2}{p_1},$$

potom klasifikujeme $\boldsymbol{x}$ *jako* π_1, *jinak klasifikujeme* $\boldsymbol{x}$ *jako* π_2.

Podobně jako v případě shodných variančních matic sestavíme výběrovou analogii klasifikačního pravidla pro případ, kdy neznáme parametry rozdělení $\boldsymbol{\mu}_i$ a $\boldsymbol{\Sigma}_i$ tak, že je nahradíme jejich příslušnými výběrovými protějšky $\bar{\boldsymbol{x}}_i$ a $\boldsymbol{S}_i$, $i = 1, 2$. Ty spočteme na základě výběru z π_1 o rozsahu n_1 a výběru z π_2 o rozsahu n_2. Kvadratická klasifikační funkce má tedy tvar

$$\begin{aligned} Q_v(\boldsymbol{x}) &= \frac{1}{2} \ln \Big(\frac{|\boldsymbol{S}_2|}{|\boldsymbol{S}_1|} \Big) - \frac{1}{2} (\bar{\boldsymbol{x}}_1' {\boldsymbol{S}_1}^{-1} \bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2' {\boldsymbol{S}_2}^{-1} \bar{\boldsymbol{x}}_2) + (\bar{\boldsymbol{x}}_1' {\boldsymbol{S}_1}^{-1} - \bar{\boldsymbol{x}}_2' {\boldsymbol{S}_2}^{-1}) \boldsymbol{x} \\ &\quad - \frac{1}{2} \boldsymbol{x}' ({\boldsymbol{S}_1}^{-1} - {\boldsymbol{S}_2}^{-1}) \boldsymbol{x} \end{aligned}$$

a získáváme *kvadratické klasifikační pravidlo.*

Pravidlo 6 *Jestliže*

$$Q_v(\boldsymbol{x}) > \ln \frac{p_2}{p_1},$$

potom klasifikujeme $\boldsymbol{x}$ *jako* π_1, *jinak jako* π_2.

Podobně jako je lineární klasifikační pravidlo 3 asymptoticky optimální v případě shodných variančních matic $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, je kvadratické klasifikační pravidlo asymptoticky optimální pro případ rozdílných variančních matic $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$. Z toho mimo jiné plyne, že v případě $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ je jednodušší lineární pravidlo lepší než složitější kvadratické pravidlo.

Až dosud jsme se v této kapitole snažili minimalizovat pravděpodobnost špatné klasifikace a nerozlišovali jsme přitom, zda byl objekt ze skupiny π_1 špatně klasifikován jako π_2 nebo naopak. V některých případech je však špatná klasifikace objektu z jedné skupiny závažnější než špatná klasifikace objektu z druhé skupiny. Například v lékařské diagnostice může mít zařazení pacienta trpícího smrtelnou chorobou mezi zdravé jedince velmi vážné následky. V takových případech se špatné klasifikace objektů z jednotlivých skupin ohodnotí různými váhami. Nechť je tedy c_1 váha špatné klasifikace objektu ze skupiny π_1 a c_2 váha špatné klasifikace objektu ze skupiny π_2. Klasifikační pravidlo 1 upravíme s ohledem na tyto váhy.

Pravidlo 7 *Jestliže*

$$c_1 p_1 f_1(\boldsymbol{x}) > c_2 p_2 f_2(\boldsymbol{x}),$$

potom klasifikujeme $\boldsymbol{x}$ *jako* π_1*, v opačném případě klasifikujeme* $\boldsymbol{x}$ *jako* π_2*.*

Váhy c_1 a c_2 často nejsou známy a jejich určení je vždy ovlivněno subjektivním postojem. Vzhledem k tomu se tento model příliš nepoužívá.

Na závěr této kapitoly zmíníme ještě jeden možný přístup ke klasifikaci. Aposteriorní pravděpodobnost skupiny π_i je podmíněná pravděpodobnost toho, že při daných hodnotách $\boldsymbol{x}$ bude pozorování $\boldsymbol{x}$ pocházet ze skupiny π_i, označme tuto pravděpodobnost $P(\pi_i|\boldsymbol{x})$, $i = 1, 2$. Z Bayesova vzorce dostáváme pro tyto pravděpodobnosti

$$\begin{aligned}
P(\pi_1|\boldsymbol{x}) &= \frac{P(\text{pozorováno } \boldsymbol{x} \text{ a zaznamenána skupina } \pi_1)}{P(\text{pozorováno } \boldsymbol{x})} \\
&= \frac{P(\text{pozorováno } \boldsymbol{x}|\pi_1)P(\pi_1)}{P(\text{pozorováno } \boldsymbol{x}|\pi_1)P(\pi_1) + P(\text{pozorováno } \boldsymbol{x}|\pi_2)P(\pi_2)} \\
&= \frac{f_1(\boldsymbol{x})p_1}{f_1(\boldsymbol{x})p_1 + f_2(\boldsymbol{x})p_2}, \\
P(\pi_2|\boldsymbol{x}) &= 1 - P(\pi_1|\boldsymbol{x}) = \frac{f_2(\boldsymbol{x})p_2}{f_1(\boldsymbol{x})p_1 + f_2(\boldsymbol{x})p_2}.
\end{aligned}$$

Protože jmenovatel ve vyjádření $P(\pi_1|\boldsymbol{x})$ a $P(\pi_2|\boldsymbol{x})$ je stejný, je použití nerovnosti (2.1) pro klasifikaci ekvivalentní s použitím nerovnosti $P(\pi_1|\boldsymbol{x}) > P(\pi_2|\boldsymbol{x})$. Můžeme tudíž klasifikační pravidlo 1 formulovat v nové podobě.

Pravidlo 8 *Jestliže*

$$P(\pi_1|\boldsymbol{x}) > P(\pi_2|\boldsymbol{x}),$$

potom klasifikujeme $\boldsymbol{x}$ *jako* π_1*, v opačném případě jako* π_2*.*

Kapitola 3

Klasifikace do většího počtu skupin

Nyní se budeme zabývat případem, kdy objekty zařazujeme do většího počtu skupin. Předpokládejme tedy, že máme k, $k > 2$ skupin objektů, označme je $\pi_1, \pi_2, \ldots, \pi_k$. Stejně jako v předchozí kapitole budeme na každém objektu pozorovat vektor p náhodných veličin $\boldsymbol{X} = (X_1, \ldots, X_p)'$. Předpokládejme, že pokud pozorování $\boldsymbol{x}$ pochází z i-té skupiny, potom má spojité p-rozměrné rozdělení s hustotu $f_i(\boldsymbol{x})$, $i = 1, 2, \ldots, k$.

3.1 Stejné varianční matice

Nejprve se zaměříme na případ, kdy varianční matice rozdělení jednotlivých skupin jsou stejné, označme $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \ldots = \boldsymbol{\Sigma}_k$. Nechť p_i je apriorní pravděpodobnost toho, že pozorování $\boldsymbol{x}$ bude pocházet ze skupiny π_i, $i = 1, 2, \ldots, k$. Optimální klasifikační pravidlo 1 přirozeně zobecníme na případ více skupin.

Pravidlo 9 *Jestliže*

$$p_i f_i(\boldsymbol{x}) \geq p_j f_j(\boldsymbol{x}) \text{ pro všechna } j = 1, 2, \ldots, k, \tag{3.1}$$

tj. jestliže $p_i f_i(\boldsymbol{x}) = \max_j p_j f_j(\boldsymbol{x})$, *potom klasifikujeme* $\boldsymbol{x}$ *jako* π_i.

Věta 4 *Nechť rozdělení ve skupinách* π_i *mají hustotu* $f_i(\boldsymbol{x})$. *Potom při použití klasifikačního pravidla 9 je minimalizována pravděpodobnost špatné klasifikace.*

Důkaz. Tato věta je důsledkem obecnější věty uvedené v knize [3], kde se navíc uvažují váhy, které zohledňují závažnost špatné klasifikace prvků z jednotlivých skupin do skupin ostatních. □
Podívejme se, jak vypadá optimální klasifikační pravidlo pro případ normálního rozdělení.

Věta 5 *Jestliže rozdělení ve skupinách π_i, $i = 1, 2, \ldots, k$, je p-rozměrné normální s parametry $(\boldsymbol{\mu_i}, \boldsymbol{\Sigma})$, potom (3.1) je ekvivalentní s nerovností*

$$\ln p_i + \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2} \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i \geq \ln p_j + \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2} \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j. \tag{3.2}$$

Důkaz. Za $f_i(\boldsymbol{x})$ ve vzorci (3.1) dosadíme hustotu normálního rozdělení (A.1), výraz $p_i f_i(\boldsymbol{x})$ zlogaritmujeme a dále upravíme. Dostáváme tak

$$\begin{aligned}
\ln[p_i f_i(\boldsymbol{x})] &= \ln p_i - \frac{1}{2} p \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i) \\
&= \ln p_i - \frac{1}{2} p \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}| \\
&\quad - \frac{1}{2} \big(\boldsymbol{x}' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \boldsymbol{x}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i + \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i \big) \\
&= \ln p_i - \frac{1}{2} p \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}| \\
&\quad - \frac{1}{2} \boldsymbol{x}' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} + \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2} \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i
\end{aligned}$$

Nerovnost (3.1) je tedy ekvivalentní s nerovností

$$\begin{aligned}
&\ln p_i - \frac{1}{2} p \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \boldsymbol{x}' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} + \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2} \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i \geq \\
&\ln p_j - \frac{1}{2} p \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \boldsymbol{x}' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} + \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2} \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j.
\end{aligned}$$

Na obou stranách nerovnice odečteme výraz $-\frac{1}{2} p \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \boldsymbol{x}' \boldsymbol{\Sigma}^{-1} \boldsymbol{x}$, který nezávisí na i, a tak dostaneme nerovnost (3.2). □
Označme

$$L_i(\boldsymbol{x}) = \ln p_i + \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2} \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i.$$

Funkce $L_i(\boldsymbol{x})$ je lineární funkcí složek vektoru $\boldsymbol{x}$, a proto se nazývá *lineární klasifikační funkce*. Můžeme tedy uvést lineární klasifikační pravidlo pro více skupin.

Pravidlo 10 *Jestliže*

$$L_i(\boldsymbol{x}) \geq L_j(\boldsymbol{x}), \text{ pro všechna } j = 1, 2, \ldots, k,$$

potom klasifikujeme $\boldsymbol{x}$ *jako* π_i.

Klasifikační pravidlo pro případ, kdy neznáme hodnoty $\boldsymbol{\mu}_i$, $i = 1, 2, \ldots, k$, a $\boldsymbol{\Sigma}$, získáme opět tak, že ve vzorci (3.2) nahradíme $\boldsymbol{\mu}_i$ a $\boldsymbol{\Sigma}$ příslušnými odhahy. Předpokládejme tedy, že pro každé $i = 1, 2, \ldots, k$ máme k dispozici náhodný výběr z π_i o rozsahu n_i, $\boldsymbol{X}_i = (\boldsymbol{X}_i^1, \boldsymbol{X}_i^2, \ldots, \boldsymbol{X}_i^{n_i})'$. Opět zdůrazněme, že $\boldsymbol{X}_i$ je matice typu $n_i \times p$. Do (3.2) dosadíme výběrové průměry $\bar{\boldsymbol{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij}$ na místo $\boldsymbol{\mu}_i$ a varianční matici $\boldsymbol{\Sigma}$ nahradíme sdruženou výběrovou maticí $\boldsymbol{S} = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1)\boldsymbol{S}_i$, kde $N = \sum_{i=1}^{k} n_k$ a $\boldsymbol{S}_i$ je výběrová varianční matice pro skupinu π_i. Dostáváme tak výběrovou lineární klasifikační funkci

$$L_{vi}(\boldsymbol{x}) = \ln p_i + \bar{\boldsymbol{x}}_i' \boldsymbol{S}^{-1} \boldsymbol{x} - \frac{1}{2}\bar{\boldsymbol{x}}_i' \boldsymbol{S}^{-1} \bar{\boldsymbol{x}}_i$$

a výběrovou variantu klasifikačního pravidla 10.

Pravidlo 11 *Jestliže*

$$L_{vi}(\boldsymbol{x}) \geq L_{vj}(\boldsymbol{x}) \text{ pro všechna } j = 1, 2, \ldots, k,$$

potom klasifikujeme pozorování $\boldsymbol{x}$ *jako* π_i.

3.2 Různé varianční matice

Pokud varianční matice $\boldsymbol{\Sigma}_i$ rozdělení jednotlivých skupin π_i, $i = 1, 2, \ldots, k$ nejsou vesměs stejné, tj. alespoň dvě z nich jsou vzájemně různé, mají klasifikační pravidla složitější podobu.

Pro případ p-rozměrného normálního rozdělení opět využijeme toho, že

$$\begin{aligned}
\ln[p_i f_i(\boldsymbol{x})] &= \ln p_i - \frac{1}{2} p \ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i) \\
&= \ln p_i - \frac{1}{2} p \ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| \\
&\quad -\frac{1}{2}\boldsymbol{\mu}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \boldsymbol{\mu}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{x}. \qquad (3.3)
\end{aligned}$$

Tak jako v případě shodných variančních matic vynecháme člen $-\frac{1}{2}p\ln(2\pi)$, který nezávisí na i a nemá tak vliv na klasifikaci. Klasifikační pravidlo 9 tak získává následující podobu.

Pravidlo 12 *Jestliže*

$$Q_i(\boldsymbol{x}) \geq Q_j(\boldsymbol{x}),\ j = 1, 2, \ldots, k,$$

kde $Q_i(\boldsymbol{x}) = \ln p_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| - \frac{1}{2}\boldsymbol{\mu}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \boldsymbol{\mu}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{x}$, *pak pozorování* $\boldsymbol{x}$ *klasifikujeme jako* π_i.

Dosazením výběrového průměru $\bar{\boldsymbol{x}}_i$ a výběrové varianční matice $\boldsymbol{S}_i$ na místo $\boldsymbol{\mu}_i$ a $\boldsymbol{\Sigma}_i$ ve vzorci (3.3) získáme *kvadratickou klasifikační funkci* příslušné skupiny

$$Q_{iv}(\boldsymbol{x}) = \ln p_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| - \frac{1}{2}\bar{\boldsymbol{x}}_i'\boldsymbol{S}_i^{-1}\bar{\boldsymbol{x}}_i + \bar{\boldsymbol{x}}_i'\boldsymbol{S}_i^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{x}'\boldsymbol{S}_i^{-1}\boldsymbol{x}.$$

Nyní můžeme uvést výběrovou analogii klasifikačního pravidla 12.

Pravidlo 13 *Jestliže*

$$Q_{iv}(\boldsymbol{x}) \geq Q_{jv}(\boldsymbol{x}), j = 1, 2, \ldots, k,$$

pak pozorování $\boldsymbol{x}$ *klasifikujeme jako* π_i.

Pro úplnost zobecníme pro případ více skupin ještě pravidlo 8, které ke klasifikaci používá aposteriorní pravděpodobnosti.

Pravidlo 14 *Jestliže*

$$P(\pi_i|\boldsymbol{x}) \geq P(\pi_j|\boldsymbol{x}),\ j = 1, 2, \ldots, k,$$

potom klasifikujeme $\boldsymbol{x}$ *jako* π_i.

Pro pravděpodobnosti $P(\pi_i|\boldsymbol{x})$ tentokrát platí

$$P(\pi_i|\boldsymbol{x}) = \frac{p_i f_i(\boldsymbol{x})}{\sum_{j=1}^{k} p_j f_j(\boldsymbol{x})}.$$

Stejně jako pro dvě skupiny i v tomto případě je jmenovatel stejný pro všechna $i = 1, 2, \ldots, k$ a pravidlo 14 je ekvivalentní s pravidlem 9.

Jak uvádí [6], některé počítačové programy počítají aposteriorní pravděpodobnosti $P(\pi_i|\boldsymbol{x})$ pro všechna pozorvání $\boldsymbol{x}_{ij}$, $i = 1, 2 \ldots, k$, $j = 1, 2, \ldots, n_i$. Princip výpočtu je založen na předpokladu normálního rozdělení. Obvyklým

postupem je nahradit parametry $\boldsymbol{\mu}_i$ a $\boldsymbol{\Sigma_i}$ ve vyjádření hustoty mnohorozměrného normálního rozdělení (A.1) příslušnými výběrovými protějšky $\bar{\boldsymbol{x}}_i$ a $\boldsymbol{S}_i$, $i = 1, 2, \ldots, k$. Dostáváme tak

$$\begin{aligned} P(\pi_i|\boldsymbol{x}) &= \frac{p_i(2\pi)^{-p/2}|\boldsymbol{S}_i|^{-1/2}e^{-(\boldsymbol{x}-\bar{\boldsymbol{x}}_i)\boldsymbol{S}_i^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_i)/2}}{\sum_{j=1}^{k} p_j(2\pi)^{-p/2}|\boldsymbol{S}_j|^{-1/2}e^{-(\boldsymbol{x}-\bar{\boldsymbol{x}}_j)\boldsymbol{S}_j^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_j)/2}} \\ &= \frac{p_i e^{-D_i^2/2}}{\sum_{j=1}^{k} p_j e^{-D_j^2/2}},\ i = 1, 2 \ldots, k, \end{aligned}$$

kde $D_i^2 = (\boldsymbol{x} - \bar{\boldsymbol{x}}_i)\boldsymbol{S}_i^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_i)$. Pro případ shodných variančních matic $\boldsymbol{\Sigma}_1 = \Sigma_2 = \cdots = \Sigma_k$ dostáváme dokonce

$$P(\pi_i|\boldsymbol{x}) = \frac{p_i e^{-Di^2/2}}{\sum_{j=1}^{k} p_j e^{-D_j^2/2}},\ i = 1, 2 \ldots, k,$$

kde pro D_i^2 tentokrát platí $D_i^2 = (\boldsymbol{x} - \bar{\boldsymbol{x}}_i)\boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_i)$ a $\boldsymbol{S}$ je nám již dobře známá sdružená výběrová varianční matice.

3.3 Odhady chyb

Ve větě 1 jsme si ukázali, že pravidlo 1 minimalizuje pravděpodobnost špatné klasifikace. V tomto odstavci se budeme touto pravděpodobností zabývat podrobněji. Většinou se omezíme na případ, kdy klasifikujeme do dvou skupin. V anglické literatuře se pro pravděpodobnost špatné klasifikace v obecném kontextu užívá termín *error rate*, v tomto textu ji pro stručnost budeme nazývat *chyba klasifikace* a budeme ji značit ER. V knize [6] je chybě klasifikace věnována celá kapitola. Některé závěry z této knihy si zde uvedeme a budeme se také držet její terminologie.

Zaměříme se na případ, kdy máme pravidla vytvořená na základě již klasifikovaných pozorování z nějakého konkrétního výběru. Bude nás zajímat pravděpodobnost špatné klasifikace nového, dosud nezařazeného objektu. Tato chyba klasifikace se značí AER (z anglického *actual error rate*) a je definována vztahem

$$\mathrm{AER} = p_1 P(\pi_2|\pi_1) + p_2 P(\pi_1|\pi_2), \tag{3.4}$$

kde p_1, p_2 jsou apriorní pravděpodobnosti a $P(\pi_i|\pi_j)$ je podobně jako v důkazu věty 1 podmíněná pravděpodobnost jevu, že dané pozorování bude

klasifikováno jako π_i, pochází-li ve skutečnosti ze skupiny π_j. Můžeme se také podívat na to, jak bude v průměru vypadat chyba klasifikace AER, pokud bychom měli k dispozici všechny možné výběry. V tomto případě platí

$$\mathsf{E}\mathrm{AER} = p_1\mathsf{E}[P(\pi_2|\pi_1)] + p_2\mathsf{E}[P(\pi_1|\pi_2)].$$

Abychom spočítali AER, potřebujeme znát parametry rozdělení dat ve skupinách. Ty však často neznáme, a proto se nyní pokusíme sestrojit odhady chyby klasifikace.

Budeme předpokládat, že rozdělení v obou skupinách jsou p-rozměrná normální s parametry $(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ a $(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$. Pro optimální klasifikační pravidlo 1 je chyba klasifikace také určena vztahem (3.4), pro tuto optimální hodnotu se však zavádí značení OER (*optimum error rate*), čili

$$\mathrm{OER} = p_1 P(\pi_2|\pi_1) + p_2 P(\pi_1|\pi_2). \tag{3.5}$$

Z pravidla 2 dostáváme

$$P(\pi_1|\pi_2) = P\big[\boldsymbol{\alpha}'\boldsymbol{x} > \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \ln\frac{p_2}{p_1}\big],$$

kde $\boldsymbol{\alpha}'\boldsymbol{x} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}$. Z věty A.3 a toho, že pozorování $\boldsymbol{x}$ pochází ve skutečnosti ze skupiny π_2, plyne, že rozdělení $\boldsymbol{\alpha}'\boldsymbol{x}$ je normální s parametry $(\boldsymbol{\alpha}'\boldsymbol{\mu}_2, \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha})$. Protože

$$\begin{aligned}\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha} &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\\ &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \Delta^2,\end{aligned}$$

kde Δ^2 je tzv. Mahalanobisova vzdálenost, je $\boldsymbol{\alpha}'\boldsymbol{x} \sim \mathsf{N}(\boldsymbol{\alpha}'\boldsymbol{\mu}_2, \Delta^2)$. Můžeme tedy $P(\pi_1|\pi_2)$ vyjádřit pomocí jednorozměrného normálního rozdělení

$$\begin{aligned}P(\pi_1|\pi_2) &= P\Big[\frac{\boldsymbol{\alpha}'\boldsymbol{x} - \boldsymbol{\alpha}'\boldsymbol{\mu}_2}{\Delta} > \frac{\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \ln(p_2/p_1) - \boldsymbol{\alpha}'\boldsymbol{\mu}_2}{\Delta}\Big]\\ &= P\Big[w > \frac{\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 - 2\boldsymbol{\mu}_2) + \ln(p_2/p_1)}{\Delta}\Big]\\ &= P\Big[w > \frac{\frac{1}{2}\Delta^2 + \ln(p_2/p_1)}{\Delta}\Big],\end{aligned}$$

kde $w = \frac{\boldsymbol{\alpha}'\boldsymbol{x} - \boldsymbol{\alpha}'\boldsymbol{\mu}_2}{\Delta}$. Využijeme symetrie distribuční funkce standardního normálního rozdělení a dostáváme

$$\begin{aligned}P(\pi_1|\pi_2) &= P\Big[w < \frac{-\frac{1}{2}\Delta^2 - \ln(p_2/p_1)}{\Delta}\Big]\\ &= \Phi\Big(\frac{-\frac{1}{2}\Delta^2 - \ln(p_2/p_1)}{\Delta}\Big).\end{aligned}$$

Podobnou úvahou dospějeme k tomu, že pro $P(\pi_2|\pi_1)$ platí

$$P(\pi_2|\pi_1) = P\big[\boldsymbol{\alpha}'\boldsymbol{x} < \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \ln\frac{p_2}{p_1}\big]$$

a $\boldsymbol{\alpha}'\boldsymbol{x} \sim \mathsf{N}(\boldsymbol{\alpha}'\boldsymbol{\mu}_1, \Delta^2)$. Analogickým postupem jako pro $P(\pi_1|\pi_2)$ získáme pro $u = \frac{\boldsymbol{\alpha}'\boldsymbol{x} - \boldsymbol{\alpha}'\boldsymbol{\mu}_1}{\Delta}$ vyjádření $P(\pi_2|\pi_1)$

$$\begin{aligned}
P(\pi_2|\pi_1) &= P\Big[\frac{\boldsymbol{\alpha}'\boldsymbol{x} - \boldsymbol{\alpha}'\boldsymbol{\mu}_1}{\Delta} < \frac{\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \ln(p_2/p_1) - \boldsymbol{\alpha}'\boldsymbol{\mu}_1}{\Delta}\Big] \\
&= P\Big[u < \frac{\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 - 2\boldsymbol{\mu}_1) + \ln(p_2/p_1)}{\Delta}\Big] \\
&= P\Big[u < \frac{-\frac{1}{2}\Delta^2 + \ln(p_2/p_1)}{\Delta}\Big] \\
&= \Phi\Big(\frac{-\frac{1}{2}\Delta^2 + \ln(p_2/p_1)}{\Delta}\Big).
\end{aligned}$$

Odtud dostáváme

$$\mathrm{OER} = p_1\Phi\Big(\frac{-\frac{1}{2}\Delta^2 + \ln(p_2/p_1)}{\Delta}\Big) + p_2\Phi\Big(\frac{-\frac{1}{2}\Delta^2 - \ln(p_2/p_1)}{\Delta}\Big). \tag{3.6}$$

Speciálně pro $p_1 = p_2 = \frac{1}{2}$ platí

$$\mathrm{OER} = \frac{1}{2}\Phi\Big(-\frac{1}{2}\Delta\Big) + \frac{1}{2}\Phi\Big(-\frac{1}{2}\Delta\Big) = \Phi\Big(-\frac{1}{2}\Delta\Big).$$

Pokud neznáme parametry $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ a $\boldsymbol{\Sigma}$, odhadneme OER tak, že v (3.6) nahradíme Mahalanobisovu vzdálenost Δ^2 veličinou

$$D^2 = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2).$$

Pro takovýto odhad OER (v anglické literatuře označovaný jako *plug-in error rate*, PER) platí

$$\mathrm{PER} = p_1\Phi\Big(-\frac{-\frac{1}{2}D^2 + \ln(p_2/p_1)}{D}\Big) + p_2\Phi\Big(-\frac{-\frac{1}{2}D^2 - \ln(p_2/p_1)}{D}\Big).$$

Pro $p_1 = p_2 = \frac{1}{2}$ platí dokonce $\mathrm{PER} = \Phi(-\frac{1}{2}D)$. Další metody pro odhad chybu klasifikace lze nalézt například v [6].

V tomto textu se zmíníme ještě o jednom neparametrickém odhadu chyby klasifikace, který získáme pomocí tzv. metody *resubstituce*. Pozorovaná data

z obou skupin zároveň uvažujeme jako jeden náhodný výběr, který náhodně rozdělíme na dvě části. V obou částech aplikujeme klasifikační pravidlo 3 na každé pozorování $\boldsymbol{x}_{1j}$, $j = 1, 2, \ldots, m_1$ a $\boldsymbol{x}_{2j}$, $j = 1, 2, \ldots, m_2$. Označíme si k_1 (resp. k_2) počet pozorování v první (resp. v druhé) části, která byla špatně klasifikována. Jako odhad chyby klasifikace vezmeme poměr špatně klasifikovaných pozorování ku celkovému počtu pozorování (*apparent error rate*, ApER), tedy

$$\text{ApER} = \frac{k_1 + k_2}{m_1 + m_2}. \tag{3.7}$$

Tuto metodu můžeme přirozeným způsobem zobecnit pro případ, kdy pozorování pocházejí z $k > 2$ skupin. Pro malé výběry má tato chyba klasifikace velký rozptyl. Protože všechna pozorování klasifikujeme do skupin podle pravidel, která jsme sestavili na jejich základě, je odhad ApER vychýlený. Ukažme si některé metody, jak se s tímto problémem vypořádat.

První z těchto metod bychom v anglické literatuře našli pod názvem *partitioning the sample*. Smysl této metody spočívá v tom, že výběr o $N = \sum_{i=1}^{k} n_k$ pozorováních ze všech skupin rozdělíme na dvě části. Na základě jedné části sestavíme klasifikační pravidla, která pak použijeme pro klasifikaci pozorování z druhé části. Odhad chyby klasifikace je v tomto případě nestranný, nevýhodou však je, že rozsah původního výběru N musí být dost velký, abychom mohli tuto metodu použít. V důsledku použití této metody bude mít také odhad chyby klasifikace větší rozptyl, než kdybychom k jeho výpočtu použili celý výběr.

Zlepšením této metody je metoda, která se v anglické literatuře označuje jako *holdout method*, nebo také *leaving-one-out method*, či *cross-validation*. Při použití této metody vezmeme $N - 1$ pozorování z celého výběru, na jejich základě vytvoříme klasifikační pravidla a aplikujeme je na zbývající pozorování. Tento postup zopakujeme pro každé pozorování. Odhad chyby, který pomocí této metody získáme, se příliš neliší od nestranného odhadu EAER.

Další odhad ApER je tzv. *bootstrap* odhad (BER), který je založený na *převýběrování (resampling)* původních výběrů. Tuto metodu si popíšeme pro dvě skupiny, kdy máme výběr ze skupiny π_1 o rozsahu n_1 a výběr z π_2 o rozsahu n_2. Z prvního výběru náhodně vybereme n_1 pozorování. Tak se stane, že některá pozorování z původního výběru budou vybrána víckrát a některá se v novém výběru vůbec neobjeví. Toto převýběrování provedeme i pro druhou skupinu. Naše klasifikační pravidla aplikujeme na staré i nově získané výběry, pro obě skupiny označíme m_{is} a m_{in}, $i = 1, 2$ počet špatně

klasifikovaných prvků ve starých a nových výběrech a spočteme

$$d_i = \frac{m_{is} - m_{in}}{n_i},\ i = 1, 2.$$

Tuto proceduru několikrát zopakujeme (v knize [6] je doporučený počet opakování 100 až 200). Získáme tak odhad

$$\mathrm{BER} = \mathrm{ApER} + \bar{d}_1 + \bar{d}_2.$$

Každá z těchto metod má své výhody a nevýhody a nelze obecně říci, která je lepší. V knize [6] je uvedena celá řada odkazů na publikace, které se touto tematikou zabývají podrobněji.

Kapitola 4

Diskriminační skóry

V dnešní době se pro náročné výpočty používají různé počítačové programy, které pomocí předdefinovaných postupů sami řeší celou řadu matematických úloh. V některých programech zaměřených na řešení statistických úloh jsou zahrnuty i metody pro řešení úloh klasifikace. Některé z nich (například program R) však k výpočtům lineárních klasifikačních pravidel používají jiné postupy, než které jsme si ukázali v předešlých kapitolách. V této kapitole se seznámíme s klasifikací založenou na diskriminačních skórech, kterou využívá program R, a ukážeme si, že pomocí této metody dospějeme ke stejným výsledkům. Tomuto tématu je v knize [6] věnována pouze krátká zmínka, proto jsem v této kapitole vycházela především z knihy [3].

Úloha klasifikace nějakého objektu do skupiny souvisí s úlohou diskriminace, tj. se situací, kdy chceme vhodně matematicky popsat jednotlivé skupiny tak, abychom je od sebe dobře rozlišili. Protože často pracujeme s mnohorozměrnými daty, transformujeme mnohorozměrné pozorování $\boldsymbol{x}$ na jednorozměrné pozorování y, se kterým se snáz pracuje a pro které získáme i přehlednější grafické výstupy. Protože hledáme co nejjednodušší transformaci $\boldsymbol{x}$, zvolíme y jako lineární kombinaci složek vektoru $\boldsymbol{x}$. Otázkou tedy je, jaké koeficienty lineární kombinace zvolit, abychom od sebe skupiny co nejlépe oddělili.

Opět předpokládejme, že máme k skupin $\pi_1, \pi_2, \ldots, \pi_k$ a na objektech pozorujeme p náhodných veličin $\boldsymbol{X} = (X_1, \ldots, X_p)'$. Nechť $\boldsymbol{\mu}_i$ je střední hodnota $\boldsymbol{X}$, pokud pozorování pochází ze skupiny π_i, a nechť $\boldsymbol{\Sigma}$ je variančí matice stejná pro všechny skupiny. Označme $\bar{\boldsymbol{\mu}} = \frac{1}{k}\sum_{i=1}^{k} \boldsymbol{\mu}_i$ vektor výběrových průměrů a $\boldsymbol{B} = \sum_{i=1}^{k} (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'$. Náhodná veličina $Y = \boldsymbol{z}'\boldsymbol{X}$, kde $\boldsymbol{z} = (z_1, z_2, \ldots, z_p)$ je vektor koeficientů lineární kombinace, má

střední hodnotu $\mu_{iY} = \mathsf{E}Y = \mathsf{E}(\boldsymbol{z}'\boldsymbol{X}) = \boldsymbol{z}'\boldsymbol{\mu}_i$, pokud pozorování pochází ze skupiny π_i. Rozptyl náhodné veličiny Y je stejný pro všechny skupiny, $\sigma_Y^2 = \mathsf{var}(\boldsymbol{z}'\boldsymbol{X}) = \boldsymbol{z}'\mathsf{var}\boldsymbol{X}\boldsymbol{z} = \boldsymbol{z}'\boldsymbol{\Sigma}\boldsymbol{z}$. Definujme průměr přes všechny skupiny

$$\bar{\mu}_Y = \frac{1}{k}\sum_{i=1}^{k}\mu_{iY} = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{z}'\boldsymbol{\mu}_i = \boldsymbol{z}'\Big(\frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\mu}_i\Big) = \boldsymbol{z}'\bar{\boldsymbol{\mu}}.$$

Zabývejme se nyní poměrem

$$\frac{\sum_{i=1}^{k}(\mu_{iY}-\bar{\mu}_Y)^2}{\sigma_Y^2} = \frac{\sum_{i=1}^{k}(\boldsymbol{z}'\boldsymbol{\mu}_i-\boldsymbol{z}'\bar{\boldsymbol{\mu}})^2}{\boldsymbol{z}'\boldsymbol{\Sigma}\boldsymbol{z}} = \frac{\boldsymbol{z}'(\sum_{i=1}^{k}(\boldsymbol{\mu}_i-\bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i-\bar{\boldsymbol{\mu}})')\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{\Sigma}\boldsymbol{z}},$$

tedy poměrem

$$\frac{\sum_{i=1}^{k}(\mu_{iY}-\bar{\mu}_Y)^2}{\sigma_Y^2} = \frac{\boldsymbol{z}'\boldsymbol{B}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{\Sigma}\boldsymbol{z}}. \tag{4.1}$$

Poměr (4.1) měří variabilitu mezi skupinami relativně vzhledem k celkové variabilitě. Protože od sebe chceme skupiny co nejlépe rozlišit, budeme hledat takový vektor koeficientů $\boldsymbol{z}$, který tento poměr maximalizuje. Pokud vektor $\boldsymbol{z}$ vynásobíme libovolnou nenulovou konstantou, hodnota (4.1) se nezmění. Je běžné vybrat takový vektor, pro který $\boldsymbol{z}'\boldsymbol{\Sigma}\boldsymbol{z} = 1$. Následující věta ukazuje, že hledaným řešením jsou koeficienty tzv. *diskriminačních skórů.*

Věta 6 *Nechť $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0$, kde $s \leq \min\{k-1, p\}$ jsou nenulová vlastní čísla matice $\boldsymbol{\Sigma}^{-1}\boldsymbol{B}$ a nechť $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_s$ jsou vlastní vektory odpovídající těmto vlastním číslům (takové, že $\boldsymbol{e}_i'\boldsymbol{\Sigma}\boldsymbol{e}_i = 1$).*
Potom vektor koeficientů, který maximalizuje poměr (4.1), je $\boldsymbol{z}_1 = \boldsymbol{e}_1$. Lineární kombinace $\boldsymbol{z}_1'\boldsymbol{X}$ se nazývá první diskriminační skór.
Mezi vektory koeficientů $\boldsymbol{z}_2$ takovými, že $\mathsf{cov}(\boldsymbol{z}_1'\boldsymbol{X}, \boldsymbol{z}_2'\boldsymbol{X}) = 0$, maximalizuje poměr (4.1) vektor $\boldsymbol{z}_2 = \boldsymbol{e}_2$. Lineární kombinace $\boldsymbol{z}_2'\boldsymbol{X}$ se nazývá druhý diskriminační skór.
Podobně vektor $\boldsymbol{z}_k = \boldsymbol{e}_k$, $k \leq s$ maximalizuje poměr (4.1) mezi všemi vektory $\boldsymbol{z}_k$ takovými, že $\mathsf{cov}(\boldsymbol{z}_k'\boldsymbol{X}, \boldsymbol{z}_i'\boldsymbol{X}) = 0$, pro $i = 1, 2, \ldots, k-1$. Lineární kombinace $\boldsymbol{z}_k'\boldsymbol{X}$ se nazývá k-tý diskriminační skór.
Navíc platí $\mathsf{var}(\boldsymbol{z}_i'\boldsymbol{X}) = 1$, $i = 1, 2, \ldots, s$.

Důkaz. Viz [3] □
Podívejme se nyní na to, proč je ve větě 6 počet diskriminačních skórů $s \leq \min\{k-1, p\}$. Popíšeme si zde odvození, které je uvedeno v [3]. Jelikož

je s zároveň počet nenulových vlastních čísel matice $\boldsymbol{\Sigma}^{-1}\boldsymbol{B}$, která je typu $p \times p$, musí být $s \leq p$. Pro vektory $\boldsymbol{\mu}_1 - \bar{\boldsymbol{\mu}}, \boldsymbol{\mu}_2 - \bar{\boldsymbol{\mu}}, \ldots, \boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}$ platí

$$(\boldsymbol{\mu}_1 - \bar{\boldsymbol{\mu}}) + (\boldsymbol{\mu}_2 - \bar{\boldsymbol{\mu}}) + \cdots + (\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}) = \sum_{i=1}^{k} \boldsymbol{\mu}_i - k\bar{\boldsymbol{\mu}} = k\bar{\boldsymbol{\mu}} - k\bar{\boldsymbol{\mu}} = \mathbf{0}.$$

Vektor $\boldsymbol{\mu}_1 - \bar{\boldsymbol{\mu}}$ lze tedy vyjádřit jako lineární kombinaci zbylých vektorů a dimenze prostoru, který tyto vektory generují je nejvýše $k-1$. Pro libovolný vektor $\boldsymbol{e}$, který je kolmý na každý vektor $\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}$, tj. pro který $(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\boldsymbol{e} = 0$, dostáváme

$$\boldsymbol{B}\boldsymbol{e} = \sum_{i=1}^{k} (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\boldsymbol{e} = \sum_{i=1}^{k} (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})0 = \mathbf{0},$$

takže $\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\boldsymbol{e} = 0\boldsymbol{e}$. Tedy existuje $p-q$ ortogonálních vektorů odpovídajících nulovému vlastnímu číslu, z čehož plyne, že existuje q nebo méně nenulových vlastních čísel matice $\boldsymbol{\Sigma}^{-1}\boldsymbol{B}$. Protože je vždy $q \leq k-1$, pro počet nenulových vlastních čísel platí $s \leq \min\{k-1, p\}$.

Nyní se pokusíme odvodit klasifikační pravidlo založené na diskriminačních skórech. Definujme vektor $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_s)'$, kde $Y_i = \boldsymbol{z}_i'\boldsymbol{X}$ je i-tý diskriminační skór z věty 6. Z této věty vidíme, že vektor Y má střední hodnotu $\boldsymbol{\mu}_{iY} = (\mu_{iY_1}, \mu_{iY_2}, \ldots, \mu_{iY_s})' = (\boldsymbol{z}_1'\boldsymbol{\mu}_i, \boldsymbol{z}_2'\boldsymbol{\mu}_i, \ldots, \boldsymbol{z}_s'\boldsymbol{\mu}_i)'$, pokud objekt pochází ze skupiny π_i, a nezávisle na tom, ze které skupiny objekt pochází, má $\boldsymbol{Y}$ jednotkovou varianční matici. Pokud pro nějaký objekt nabývá $\boldsymbol{X}$ hodnoty $\boldsymbol{x}$, označme $\boldsymbol{y} = (y_1, y_2, \ldots, y_s) = (\boldsymbol{z}_1'\boldsymbol{x}, \boldsymbol{z}_2'\boldsymbol{x}, \ldots, \boldsymbol{z}_s'\boldsymbol{x})'$.

Protože složky $\boldsymbol{Y}$ mají jednotkový rozptyl a nulovou kovarianci, dostáváme pro čtverec vzdálenosti $\boldsymbol{y}$ a $\boldsymbol{\mu}_i$ v jednotlivých skupinách

$$(\boldsymbol{y} - \boldsymbol{\mu}_{iY})(\boldsymbol{y} - \boldsymbol{\mu}_{iY})' = \sum_{j=1}^{s} (y_j - \mu_{iY_j})^2.$$

Zdá se rozumné zařadit pozorování $\boldsymbol{x}$ do té skupiny π_i, pro kterou je vzdálenost $\boldsymbol{y}$ od $\boldsymbol{\mu}_{iY}$ nejmenší.

Pravidlo 15 *Jestliže*

$$\sum_{j=1}^{s} (y_j - \mu_{iY_j})^2 = \sum_{j=1}^{s} [\boldsymbol{z}_j'(\boldsymbol{x} - \boldsymbol{\mu}_i)]^2 \leq \sum_{j=1}^{s} [\boldsymbol{z}_j'(\boldsymbol{x} - \boldsymbol{\mu}_l)]^2 \quad \textit{pro } l = 1, 2, \ldots, k,$$

potom klasifikujeme $\boldsymbol{x}$ *jako* π_i.

Nyní si ukážeme, že pro apriorní pravděpodobnosti $p_1 = p_2 = \cdots = p_k = \frac{1}{k}$ je pravidlo 15 ekvivalentní s pravidlem 10, které ke klasifikaci využívá linearní klasifikační funkce $L_i(\boldsymbol{x}) = \ln p_i + \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2} \boldsymbol{\mu}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i$.

Věta 7 *Nechť $y_j = \boldsymbol{z}_j' \boldsymbol{x}$, kde $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j$ a $\boldsymbol{e}_j$ je vlastní vektor matice $\boldsymbol{\Sigma}^{-1/2} \boldsymbol{B} \boldsymbol{\Sigma}^{-1/2}$. Potom*

$$\begin{aligned} \sum_{j=1}^{p} (y_j - \mu_{iY_j})^2 &= \sum_{j=1}^{p} [\boldsymbol{z}_j' (\boldsymbol{x} - \boldsymbol{\mu}_i)]^2 \\ &= (\boldsymbol{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i) - L_i(\boldsymbol{x}) + \frac{1}{2} \boldsymbol{x}' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} + \ln p_i. \end{aligned}$$

Jestliže $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0 = \lambda_{s+1} = \cdots = \lambda_p$, potom je výraz $\sum_{j=s+1}^{p} (y_j - \mu_{iY_j})^2$ konstantní pro všechny skupiny $\pi_1, \pi_2, \ldots, \pi_k$, a pouze $y_1, y_2, \ldots, y_s$, a tedy $\sum_{j=1}^{s} (y_j - \mu_{iY_j})^2$, $i = 1, 2, \ldots, k$, má vliv na klasifikaci.

Důkaz. Viz [3] □

Z věty 7 vidíme, že index i, který minimalizuje výraz $\sum_{j=1}^{p} (y_j - \mu_{iY_j})^2$ v pravidle 15, zároveň maximalizuje výraz $L_i(\boldsymbol{x}) - \frac{1}{2} \boldsymbol{x}' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \ln p_i$. Pro $p_1 = p_2 = \cdots = p_k = \frac{1}{k}$ je $-\frac{1}{2} \boldsymbol{x}' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \ln p_i = -\frac{1}{2} \boldsymbol{x}' \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \ln \frac{1}{k}$ konstantní, a tudíž pro i nabývá $L_i(\boldsymbol{x})$ maxima. Odtud dostáváme, že pro $p_1 = p_2 = \cdots = p_k = \frac{1}{k}$ je použití pravidel 15 a 10 ekvivalentní.

Z věty 7 také vyplývá možná výhoda použití diskriminačních skórů ke klasifikaci. Zatímco pro linární klasifikační pravidlo je potřeba pro každé pozorování spočítat p hodnot lineárních klasifikačních funkcí, diskriminačních skórů stačí spočítat pouze s, kde $s \leq \min\{k-1, p\}$. Počet skupin, do kterých klasifikujeme, je často mnohem menší než počet veličin, které na objektech pozorujeme. V těchto případech pak může být vyčíslení diskriminačních skórů výpočetně méně náročné.

Pokud neznáme hodnoty $\boldsymbol{\mu}_i$ a $\boldsymbol{\Sigma}$, potom podobně jako při klasifikaci použijeme místo nich příslušné odhady. Předpokládejme tedy, že pro $i = 1, 2, \ldots, k$ máme náhodný výběr $\boldsymbol{X}_i = (\boldsymbol{X}_i^1, \boldsymbol{X}_i^2, \ldots, \boldsymbol{X}_i^{n_i})'$ z rozdělení skupiny π_i. Stejně jako v předchozích kapitolách spočteme vektor výběrových průměrů $\bar{\boldsymbol{x}}_i$ a výběrové varianční matice $\boldsymbol{S}_i$. Dále definujme vektor průměrů přes všechny skupiny $\bar{\boldsymbol{x}} = \frac{1}{N} \sum_{i=1}^{k} n_i \bar{\boldsymbol{x}}_i = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}$, kde $N = \sum_{i=1}^{k} n_i$, a matici mezitřídní variability

$$\hat{\boldsymbol{B}} = \sum_{i=1}^{k} (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})',$$

výběrový protějšek matice $\boldsymbol{B}$. Definujme také matici vnitrotřídní variability

$$\boldsymbol{W} = \sum_{i=1}^{k} (n_i - 1)\boldsymbol{S}_i = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'.$$

Tedy $\frac{1}{N-k}\boldsymbol{W} = \boldsymbol{S}$ je také odhad $\boldsymbol{\Sigma}$. Protože $N - k$ je konstanta, nabývá poměr $\hat{\boldsymbol{z}}'\hat{\boldsymbol{B}}\hat{\boldsymbol{z}}/\hat{\boldsymbol{z}}'\boldsymbol{S}\hat{\boldsymbol{z}}$ maxima pro stejný vektor $\boldsymbol{z}$ jako poměr $\hat{\boldsymbol{z}}'\hat{\boldsymbol{B}}\hat{\boldsymbol{z}}/\hat{\boldsymbol{z}}'\boldsymbol{W}\hat{\boldsymbol{z}}$. Jak je uvedeno v knize [3], je běžné převést úlohu maximalizace vektoru $\hat{\boldsymbol{z}}$ na hledání vlastních vektorů $\hat{\boldsymbol{e}}_i$ matice $\boldsymbol{W}^{-1}\hat{\boldsymbol{B}}$, neboť je-li $\boldsymbol{W}^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{e}} = \hat{\lambda}\hat{\boldsymbol{e}}$, potom $\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{e}} = \hat{\lambda}(N-k)\hat{\boldsymbol{e}}$. Následující věta (výběrová analogie věty 6) je bez důkazu uvedena v knize [3].

Věta 8 *Nechť $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_s > 0$, kde $s \leq \min\{k-1, p\}$ jsou nenulová vlastní čísla matice $\boldsymbol{W}^{-1}\hat{\boldsymbol{B}}$ a nechť $\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \ldots, \hat{\boldsymbol{e}}_s$ jsou vlastní vektory odpovídající těmto vlastním číslům (takové, že $\hat{\boldsymbol{e}}_i'\boldsymbol{\Sigma}\hat{\boldsymbol{e}}_i = 1$).*
Potom vektor koeficientů $\hat{\boldsymbol{z}}$, který maximalizuje poměr

$$\frac{\hat{\boldsymbol{z}}'\hat{\boldsymbol{B}}\hat{\boldsymbol{z}}}{\hat{\boldsymbol{z}}'\boldsymbol{W}\hat{\boldsymbol{z}}} = \frac{\hat{\boldsymbol{z}}'[\sum_{i=1}^{k}(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})']\hat{\boldsymbol{z}}}{\hat{\boldsymbol{z}}'[\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)']\hat{\boldsymbol{z}}},$$

je $\hat{\boldsymbol{z}}_1 = \hat{\boldsymbol{e}}_1$. Lineární kombinace $\hat{\boldsymbol{z}}_1'\boldsymbol{x}$ se nazývá první výběrový diskriminační skór. Pro volbu vektoru koeficientů $\hat{\boldsymbol{z}}_2 = \hat{\boldsymbol{e}}_2$ se lineární kombinace $\hat{\boldsymbol{z}}_2'\boldsymbol{x}$ nazývá druhý výběrový diskriminační skór. Analogicky $\hat{\boldsymbol{z}}_2'\boldsymbol{x} = \hat{\boldsymbol{e}}_2'\boldsymbol{x}$ se nazývá k-tý výběrový diskriminační skór, $k \leq s$.

Narozdíl od věty 6 nemají obecně výběrové diskriminační skóry nulové kovariance, spíše bývá splněna podmínka

$$\hat{\boldsymbol{z}}_i'\boldsymbol{S}\hat{\boldsymbol{z}}_k = \begin{cases} 1 & \text{jestliže } i = k \leq s \\ 0 & \text{jinak} \end{cases}$$

Na závěr zformulujeme výběrovou analogii klasifikačního pravidla 15.

Pravidlo 16 *Jestliže*

$$\sum_{j=1}^{s}(y_j - \hat{y}_{kj})^2 = \sum_{j=1}^{s}[\hat{\boldsymbol{z}}_j'(\boldsymbol{x} - \hat{\boldsymbol{x}}_k)]^2 \leq \sum_{j=1}^{s}[\hat{\boldsymbol{z}}_j'(\boldsymbol{x} - \hat{\boldsymbol{x}}_i)]^2 \quad \textit{pro } i = 1, 2, \ldots, k,$$

potom klasifikujeme $\boldsymbol{x}$ jako π_k.

Pravidlo 16 je pro $p_1 = p_2 = \cdots = p_k = \frac{1}{k}$ ekvivalentní s pravidlem 11, s využitím věty 8 a stejné argumentace jako v případě známých parametrů.

Diskriminační skóry budou spočítány pro konkrétní data v kapitole 6.

Kapitola 5

Logistická klasifikace

V této kapitole jsem vycházela převážně z knihy [2]. V předchozích kapitolách jsme sestavovali klasifikační pravidla pro vektor pozorování, který pocházel ze spojitého rozdělení s hustotou $f_i(x)$. V této kapitole se omezíme na případ, kdy klasifikujeme pouze do dvou skupin, jejichž rozdělení mají stejnou kovarianční matici. Pro mnohorozměrné normální rozdělení je klasifikační pravidlo 2 založené na poměru

$$\begin{aligned} \ln\left[\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}\right] &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \\ &= \alpha + \boldsymbol{\beta}'\boldsymbol{x}. \end{aligned} \tag{5.1}$$

Také pro některá jiná rozdělení je $\ln(f_1(\boldsymbol{x})/f_2(\boldsymbol{x}))$ lineární funkcí $\boldsymbol{x}$. K podobnému výsledku můžeme dospět i pro některá diskrétní rozdělení, pokud hustoty rozdělení nahradíme příslušnými pravděpodobnostmi. Můžeme tak sestavit obecnější *logistické klasifikační pravidlo*, které nevyžaduje předpoklad normálního rozdělení a používá se i v případech, kdy na objektech pozorujeme diskrétní veličiny.

Pravidlo 17 *Jestliže*

$$\alpha + \boldsymbol{\beta}'\boldsymbol{x} > \ln\frac{p_2}{p_1},$$

potom klasifikujeme $\boldsymbol{x}$ *jako* π_1*, jinak jako* π_2*.*

Logistická klasifikace je považována za obecnou metodu pro obecná rozdělení f_1 a f_2, je však třeba zdůraznit, že $\ln(f_1(\boldsymbol{x})/f_2(\boldsymbol{x}))$ nemusí být lineární funkcí $\boldsymbol{x}$, a tedy logistická klasifikace může být nevhodná metoda pro konkrétní f_1

a f_2. Pro taková rozdělení, pro která $\ln(f_1(\boldsymbol{x})/f_2(\boldsymbol{x})) = \alpha + \boldsymbol{\beta}'\boldsymbol{x}$ odvodíme tvar aposteriorních pravděpodobností $P(\pi_i|\boldsymbol{x})$, $1, 2$

$$\begin{aligned} P(\pi_1|\boldsymbol{x}) &= \frac{p_1 f_1(\boldsymbol{x})}{p_1 f_1(\boldsymbol{x}) + p_2 f_2(\boldsymbol{x})} = \frac{\frac{p_1 f_1(\boldsymbol{x})}{p_2 f_2(\boldsymbol{x})}}{\frac{p_1 f_1(\boldsymbol{x})}{p_2 f_2(\boldsymbol{x})} + 1} = \frac{\exp\{\ln \frac{p_1 f_1(\boldsymbol{x})}{p_2 f_2(\boldsymbol{x})}\}}{\exp\{\ln \frac{p_1 f_1(\boldsymbol{x})}{p_2 f_2(\boldsymbol{x})}\} + 1} \\ &= \frac{\exp\{\ln \frac{p_1}{p_2} + \ln \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}\}}{\exp\{\ln \frac{p_1}{p_2} + \ln \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}\} + 1} = \frac{e^{\ln(p_1/p_2)+\alpha+\boldsymbol{\beta}'\boldsymbol{x}}}{1 + e^{\ln(p_1/p_2)+\alpha+\boldsymbol{\beta}'\boldsymbol{x}}} \\ &= \frac{e^{\beta_0+\boldsymbol{\beta}'\boldsymbol{x}}}{1 + e^{\beta_0+\boldsymbol{\beta}'\boldsymbol{x}}}, \quad \text{kde } \beta_0 = \ln \frac{p_1}{p_2} + \alpha. \end{aligned}$$

Pro $P(\pi_2|\boldsymbol{x})$ platí

$$P(\pi_2|\boldsymbol{x}) = 1 - P(\pi_1|\boldsymbol{x}) = 1 - \frac{e^{\beta_0+\boldsymbol{\beta}'\boldsymbol{x}}}{1 + e^{\beta_0+\boldsymbol{\beta}'\boldsymbol{x}}} = \frac{1}{1 + e^{\beta_0+\boldsymbol{\beta}'\boldsymbol{x}}}.$$

Parametry α a $\boldsymbol{\beta}$ odhadneme pomocí logistické regrese. Abychom zdůraznili, že hodnoty $P(\pi_i|\boldsymbol{x})$ závisí na parametrech β_0 a $\boldsymbol{\beta}$, označme $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta}')'$ a $p(\boldsymbol{x}; \theta) = P(\pi_1|\boldsymbol{x})$. Je tedy $P(\pi_2|\boldsymbol{x}) = 1 - p(\boldsymbol{x}; \boldsymbol{\theta})$. Jako odhad vektoru parametrů $\boldsymbol{\theta}$ použijeme maximálně věrohodný odhad. Mějme N pozorování z obou skupin a pro každé pozorování $\boldsymbol{x}_j$, $j = 1, 2, \ldots, N$ definujme náhodnou veličinu y_j, která nabývá hodnoty 1, pokud $\boldsymbol{x}_j$ pochází ze skupiny π_1, a hodnoty 0, pokud $\boldsymbol{x}_j$ pochází ze skupiny π_2. Pro zjednodušení zápisu označme ještě $\boldsymbol{z}_j = (1, \boldsymbol{x}_j')'$. Protože $\boldsymbol{\theta}'\boldsymbol{z}_j = \beta_0 + \boldsymbol{\beta}'\boldsymbol{x}$, můžeme bez újmy na obecnosti změnit značení $p(\boldsymbol{z}_j; \boldsymbol{\theta}) = p(\boldsymbol{x}_j; \boldsymbol{\theta})$. Logaritmická věrohodnostní funkce má tvar

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{j=1}^{N} \{y_j \ln p(\boldsymbol{z}_j; \boldsymbol{\theta}) + (1 - y_j) \ln(1 - p(\boldsymbol{z}_j; \boldsymbol{\theta})\} \\ &= \sum_{j=1}^{N} \{y_j \ln \frac{p(\boldsymbol{z}_j; \boldsymbol{\theta})}{1 - p(\boldsymbol{z}_j; \boldsymbol{\theta})} + \ln(1 - p(\boldsymbol{z}_j; \boldsymbol{\theta}))\} \\ &= \sum_{j=1}^{N} \{y_j \ln \frac{e^{\boldsymbol{\theta}'\boldsymbol{z}_j}/(1 + e^{\boldsymbol{\theta}'\boldsymbol{z}_j})}{1/(1 + e^{\boldsymbol{\theta}'\boldsymbol{z}_j})} + \ln \frac{1}{1 + e^{\boldsymbol{\theta}'\boldsymbol{z}_j}}\} \\ &= \sum_{j=1}^{N} \{y_j \boldsymbol{\theta}'\boldsymbol{z}_j - \ln(1 + e^{\boldsymbol{\theta}'\boldsymbol{z}_j})\}. \end{aligned}$$

Abychom našli maximum logaritmické věrohodnostní funkce $l(\boldsymbol{\theta})$, položíme její derivaci rovnou nule

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{N} \boldsymbol{z}_j [y_j - p(\boldsymbol{z}_j; \boldsymbol{\theta})] = 0 \tag{5.2}$$

a tím dostáváme soustavu $p+1$ nelineárních rovnic v proměnné $\boldsymbol{\theta}$. Protože $z_{j1} = 1$ pro všechna $j = 1, 2, \ldots, N$, dostáváme z první nerovnosti $\sum_{j=1}^{N} y_j = \sum_{j=1}^{N} p(\boldsymbol{z}_j; \boldsymbol{\theta})$. Tato rovnice vyjadřuje stav, kdy se střední hodnota počtu prvků v každé skupině rovná jejich skutečnému počtu. Uvedeme si zde Newtonův-Raphsonův iterační algoritmus pro řešení soustavy rovnic (5.2), tak jak je uveden v knize [2]. Tento algoritmus při výpočtech pracuje s druhými derivacemi logaritmické věrohodnostní funkce,

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\sum_{j=1}^{N} \boldsymbol{z}_j \boldsymbol{z}_j' p(\boldsymbol{z}_j; \boldsymbol{\theta})(1 - p(\boldsymbol{z}_j; \boldsymbol{\theta})).$$

V každém kroku algoritmu vezmeme hodnotu $\boldsymbol{\theta}_m$, kterou jsme vypočítali v minulém kroku, a spočteme novou hodnotu $\boldsymbol{\theta}_{m+1}$ podle vzorce

$$\boldsymbol{\theta}_{m+1} = \boldsymbol{\theta}_m - \Big(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\Big)^{-1} \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \tag{5.3}$$

kde derivace jsou vyčísleny pro $\boldsymbol{\theta}_m$. Běžně se užívá maticového zápisu této rovnice. Označme $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)'$, $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N)'$. Zdůrazněme, že $\boldsymbol{Z}$ je matice typu $N \times (p+1)$, jejíž řádky tvoří vektory $\boldsymbol{z}_j$. Dále označme $\boldsymbol{p} = (p(\boldsymbol{z}_1; \boldsymbol{\theta}), p(\boldsymbol{z}_2; \boldsymbol{\theta}), \ldots, p(\boldsymbol{z}_N; \boldsymbol{\theta}))'$ a $\boldsymbol{W}$ diagonální matici, jejíž i-tý prvek na diagonále je $p(\boldsymbol{z}_i; \boldsymbol{\theta})(1 - p(\boldsymbol{z}_i; \boldsymbol{\theta}))$. Potom $\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{Z}'(\boldsymbol{y} - \boldsymbol{p})$ a $\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z}$. V maticovém zápisu má tedy (5.3) tvar

$$\begin{aligned} \boldsymbol{\theta}_{m+1} &= \boldsymbol{\theta}_m + (\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z})^{-1}\boldsymbol{Z}'(\boldsymbol{y} - \boldsymbol{p}) \\ &= (\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{W}(\boldsymbol{Z}\boldsymbol{\theta}_m + \boldsymbol{W}^{-1}(\boldsymbol{y} - \boldsymbol{p})) \\ &= (\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{a}, \end{aligned}$$

kde $\boldsymbol{a} = \boldsymbol{Z}\boldsymbol{\theta}_m + \boldsymbol{W}^{-1}(\boldsymbol{y} - \boldsymbol{p})$. Tento krok opakujeme tak dlouho, dokud se mění $\boldsymbol{p}$, a tedy i $\boldsymbol{W}$ a $\boldsymbol{a}$. Často se jako počáteční hodnota $\boldsymbol{\theta}_m$ v prvním kroku algoritmu volí $\boldsymbol{\theta}_0 = 0$. Konvergence algoritmu zaručena není, ale ve většině případů algoritmus řešení najde.

Kapitola 6

Příklad

V této kapitole si předvedeme použití některých klasifikačních pravidel v praxi. V některých knihách (např. [6] a [3]), jsou uvedeny příklady na použití lineárních klasifikačních pravidel, soubory dat jsou však příliš malé na to, abychom si na nich předvedli i použití kvadratických klasifikačních pravidel. Při kvadratické klasifikaci se totiž odhaduje větší množství parametrů, neboť předpokládáme, že varianční matice jsou v jednotlivých skupinách různé. Proto jsem se rozhodla použít jiná data, která sice nedávají tak dobré výsledky, ale odpovídají reálné situaci.

Data obsahují vybrané údaje o smrkových porostech z několika lokalit v České republice. Pro každý porost máme zaznamemnány hodnoty veličin plocha, vek, zakme, zast, tlous, vyska, bonita, zastab, zassku a vyczak. Popišme si, co jednotlivé veličiny znamenají. Veličiny plocha, vek a vyska obsahují údaje o ploše porostu (v hektarech), jeho střední výšce (v metrech) a stáří, tlous udává průměrnou tloušťku kmene (v centimetrech) ve výšce 130 cm nad zemí v daném porostu. Veličina zast udává procentuální zastoupení smrku mezi ostatními dřevinami v daném porostu. Proměnné zassku a zastab udávají skutečnou a tabulkovou zásobu dřeva (v m^3/ha), tj. kolik dřeva je v daném porostu na ploše jednoho hektaru. Veličina zakme udává zakmenění, které lze vyjádřit jako poměr skutečné a tabulkové zásoby dřeva. Proměnná vyczak udává výčetní kruhovou základnu, která představuje součet ploch průřezů kmenů všech stromů ve výšce 130 cm v porostu o výměře jeden hektar. Proměnná bonita vyjadřuje produkční schopnost porostu, a proto je v lesnictví důležitým ukazatelem kvality porostu. V našem případě jsou použity tzv. absolutní výškové bonity, které udávají střední výšku, jakou by měl mít daný porost ve sto letech. Bonita závisí především na věku a

současné výšce daného porostu, její hodnoty jsou odstupňovány po dvou metrech a v praxi se získávají z tabulek. Bližší informace lze nalézt např. v [5] nebo v základních učebnicích dendrometrie.

Naším cílem bude klasifikace porostů do skupin se stejnou bonitou. Abychom si předvedli i logistickou klasifikaci, které byla věnována kapitola 5, budeme pracovat pouze s dvěma skupinami porostů s bonitami 26 a 28, které se v lesnické praxi i v našem datovém souboru objevují nejčastěji. Ostatně pro klasifikaci do většího počtu skupin se naše data nehodí, neboť chyba klasifikace je v tomto případě velká (kolem 40% pro čtyři skupiny).

K dispozici máme $N = 1209$ pozorování, přičemž pro $n_1 = 538$ z nich byla tabulkou zjištěna bonita 26 a zbylých $n_2 = 671$ pozorování má bonitu 28. Budeme zkoumat chybu klasifikace ApER (3.7) postupně pro lineární, kvadratickou a logistickou klasifikaci a porovnáme také odhady této chyby klasifikace při použití metod uvedených v odstavci 3.3. Tyto odhady budeme také značit ApER s tím, že z kontextu bude jasné, o jaké odhady se jedná.

Pro výpočty jsem použila program R, který je volně šiřitelný. Pro lineární klasifikaci je v tomto programu definovaná funkce lda. Funkce lda má několik volitelných parametrů, jedním z nich je nastavení apriorních pravděpodobností p_1 a p_2. Původní nastavení funkce lda je takové, že apriorní pravděpodobnosti jsou odhadnuty relativními četnostmi tak, jak bylo popsáno v odstavci 2.2. Protože vzorek našich dat byl získán jen v určitých lokalitách a neodpovídá tedy stavu na celém území České republiky, změníme hodnotu apriorních pravděpodobností na $p_1 = p_2 = \frac{1}{2}$. Pro toto nastavení dostáváme chybu klasifikace

$$\text{ApER} = 0.2390405\,.$$

Jak už jsme zmínili v kapitole 4, program R využívá ke klasifikaci diskriminačních skórů. Připomeňme, že počet diskriminačních skórů je vždy nejvýše $k - 1$, kde k je počet skupin. V našem případě tak program pracuje pouze s prvním diskriminačním skórem $Y_1 = \boldsymbol{z}_1'\boldsymbol{X}$. Jako výstup lda získáme i koeficienty $\boldsymbol{z}_1$. Ty jsou uvedny v tabulce 6.1. Z této tabulky vidíme, že v absolutní hodnotě je největší koeficient u proměnné vyska. Mohli bychom se proto domnívat, že střední výška porostu nejvíce ovlivní diskriminaci a tím tedy i klasifikaci. Tato domněnka však není správná, neboť veličiny mají různé jednotky.

Diskriminačního skóru Y využijeme také jako transformace našich devítirozměrných pozorování $\boldsymbol{x}_{i1}, \boldsymbol{x}_{j2}$, $i = 1, 2, \ldots, n_1$, $j = 1, 2, \ldots, n_2$, na jednorozměrná pozorování y_{i1}, y_{j2}, pro která získáme přehlednější grafické výstupy. Jak vidíme z histogramů v obrázku 6.1, hodnoty Y jsou pro skupinu

Tabulka 6.1: Koeficienty diskriminačního skóru

	LD1
plocha	-0.023772342
vek	-0.134172393
zakme	0.304551981
zast	0.012997966
tlous	0.033078305
vyska	-2.425319266
zastab	0.124178542
zassku	-0.003884721
vyczak	-0.009585625

porostů s bonitou 28 posunuty více doprava a obě skupiny tak od sebe můžeme dobře rozeznat.

Pro odhad ApER použijeme nejprve metodu *partitioning the sample*, která je popsána v odstavci 3.3. Touto metodou dostaneme odhad chyby klasifikace

$$\text{ApER} = 0.261157\,.$$

ApER je v tomto případě o něco větší, což odpovídá tomu, že klasifikační pravidla byla založena na výběru o polovičním rozsahu.

Pokud nastavíme parametr CV funkce lda na hodnotu TRUE, získáme i odhad ApER metodou *holdout*. Odhad ApER je v tomto případě

$$\text{ApER} = 0.2415219\,.$$

Pro kvadratickou klasifikaci je v programu R definována funkce qda. Stejně jako v případě lineární klasifikace nastavíme hodnoty apriorních pravděpodobností na $p_1 = p_2 = \frac{1}{2}$. Chyba klasifikace ApER je tentokrát

$$\text{ApER} = 0.2109181.$$

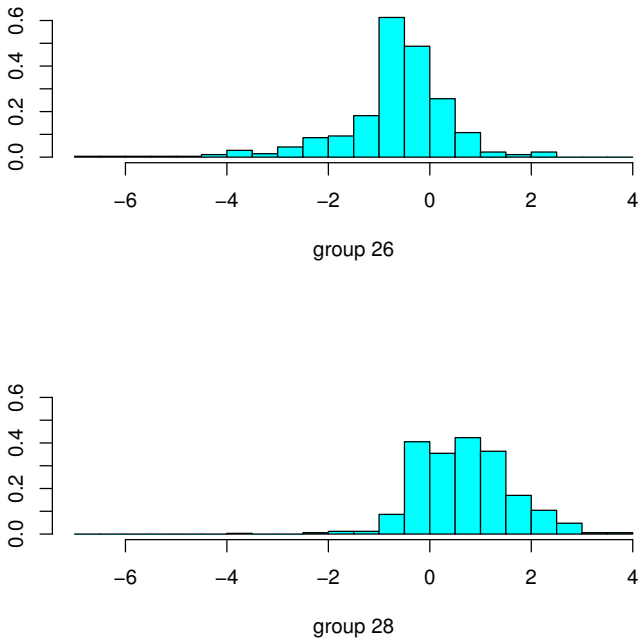
Opět odhadneme tuto chybu nejprve pomocí metody *partitioning the sample*. Dostáváme tak

$$\text{ApER} = 0.2516556.$$

Pro *holdout* metodu dostaneme odhad

$$\text{ApER} = 0.2291150.$$

Obrázek 6.1: Histogramy diskriminačních skórů pro obě skupiny



Zabývejme se nyní modelem logistické klasifikace (5.1). Připomeňme, že se snažíme odhadnout koeficienty $\boldsymbol{\theta} = (\beta_0, \beta_1, \ldots, \beta_p)$, kde $p = 9$ je počet proměnných. V programu R k tomu účelu použijeme definovanou funkci glm, jejíž parametr family nastavíme na binomial. Dostaneme tak odhady které jsou uvedené v tabulce 6.2.

Pro tyto hodnoty koeficientů dostáváme chybu klasifikace

$$\text{ApER} = 0.2084367.$$

Pro odhad každého parametru β_i, $i = 0, 1, \ldots, 10$ obsahuje tabulka 6.2 také hodnoty *z value*, které odpovídají testu nulové hypotézy, že koeficient u dané proměnné je nulový, zatímco u ostatních proměnných je nenulový. Jak dále z tabulky 6.2 zjistíme, tuto hypotézu nezamítáme na hladině 5% hned pro několik proměnných (plocha, zast, tlous, zassku, vyczak). Protože pro

Tabulka 6.2: Odhady koeficientů v modelu logistické klasifikace

	Estimate	Std. Error	z value	p-value	
(Intercept)	18.466686	3.134753	5.891	3.84e-09	***
plocha	-0.051765	0.029823	-1.736	0.08261	.
vek	-0.259603	0.017581	-14.766	< 2e-16	***
zakme	0.510472	0.196659	2.596	0.00944	**
zast	0.023922	0.019188	1.247	0.21250	
tlous	-0.001361	0.053342	-0.026	0.97964	
vyska	-4.095917	0.302333	-13.548	< 2e-16	***
zastab	0.217833	0.015316	14.222	< 2e-16	***
zassku	-0.005059	0.004528	-1.117	0.26383	
vyczak	-0.040665	0.087223	-0.466	0.64106	

proměnnou tlous dostáváme nejvyšší p-hodnotu, vyřadíme ji z našeho modelu. Celý proces zopakujeme, tentokrát pro zbylých osm proměnných. Jelikož nám v tomto případě vyjde nejvyšší p-hodnota (0.58145) pro proměnnou vyczak, vyloučíme ji z našeho modelu. Takto postupujeme tak dlouho, dokud u nějaké proměnné nezamítáme nulovou hypotézu na hladině 5%. Tento postup je doporučený v knize [2]. Ve výsledném modelu nám zbydou proměnné vek, zakme, zast, vyska, zastab a zassku. Teprve tento zúžený model považujeme za vhodný a odhady příslušných parametrů najdeme v tabulce 6.3.

Tabulka 6.3: Odhady koeficientů v zúženém modelu logistické klasifikace

	Estimate	Std. Error	z value	p-value	
(Intercept)	19.036382	1.988593	9.573	< 2e-16	***
vek	-0.255944	0.017298	-14.796	< 2e-16	***
zakme	0.447527	0.097893	4.572	4.84e-06	***
zast	0.015302	0.006136	2.494	0.012639	*
vyska	-4.103360	0.281459	-14.579	< 2e-16	***
zastab	0.217191	0.014226	15.267	< 2e-16	***
zassku	-0.006922	0.002047	-3.382	0.000719	***

Pro tento model a odhady koeficientů z tabulky 6.3 dostáváme chybu klasifikace

$$\text{ApER} = 0.2125724.$$

Vidíme, že chyba klasifikace je i pro tento zúžený logistický model menší než chyba (resp. různé odhady chyby) při použití lineární a kvadratické klasifikace. Přitom je jen nepatrně horší než ApER pro původní model logistické klasifikace.

Výběrová lineární i kvadratická klasifikační pravidla předpokládají normální rozdělení dat. Naše data tento předpoklad nesplňují, přesto byly výsledky lineární a kvadratické klasifikace srovnatelné s výsledky logistické klasifikace, která normalitu nepředpokládá. Ve všech případech se zdá být chyba klasifikace poměrně vysoká - přibližně pětina až čtvrtina pozorování byla chybně klasifikována. To si můžeme vysvětlit tím, že bonita porostu je tabulková hodnota odstupňovaná po dvou metrech. Může se tak například stát, že porosty stejného stáří, jejichž střední výška se málo liší, mohou mít podle tabulek odlišné bonity. Navíc většina veličin, které na porostech sledujeme, by se v praxi určovala přesně jen velmi těžko, a proto se používá různých osvědčených metod pro jejich odhad. Samotná data tak mohou obsahovat nepřesnosti, které se následně projeví při klasifikaci. Vzhledem k tomu můžeme považovat výsledky s čtvrtinovou chybou za poměrně úspěšné.

Dodatek

Zde připomeneme některé pojmy a uvedeme některé věty, které se v textu používají.

Definice A.1 *Nechť* $\boldsymbol{X} = (X_1, \ldots, X_p)'$ *je náhodný vektor,* $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ *je daný vektor a* $\boldsymbol{\Sigma} = (\sigma_{ij})$ *je symetrická pozitivně semidefinitní matice typu* $p \times p$. *Řekneme, že* $\boldsymbol{X}$ *má p-rozměrné normální rozdělení s parametry* $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *jestliže pro libovolný vektor* $\boldsymbol{c} \in \mathbb{R}_p$ *platí*

$$\boldsymbol{c'X} \sim \mathsf{N}(\boldsymbol{c'\mu}, \boldsymbol{c'\Sigma c}).$$

Je-li $\boldsymbol{\Sigma}$ *regulární, mluvíme o regulárním p-rozměrném rozdělení. Je-li* $\boldsymbol{\Sigma}$ *singulární, jde o singulární p-rozměrné rozdělení. Jestliže* $\boldsymbol{X}$ *má p-rozměrné normální rozdělení s parametry* $(\boldsymbol{\mu},\boldsymbol{\Sigma})$, *značíme to jako* $\boldsymbol{X} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *případně podrobněji jako* $\boldsymbol{X} \sim \mathsf{N_p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Věta A.1 *Je-li* $\boldsymbol{X} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *pak* $\mathsf{E}\boldsymbol{X} = \boldsymbol{\mu}$, $\mathsf{var}\boldsymbol{X} = \boldsymbol{\Sigma}$.

Důkaz. Viz [1]. □

Věta A.2 *Nechť* $\boldsymbol{X} = (X_1, \ldots, X_p)'$ *má regulární normální rozdělení s parametry* $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Pak existuje jeho hustota* f *a je dána vzorcem*

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\Big\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\Big\}. \tag{A.1}$$

Důkaz. Viz [1]. □

V celém tomto textu pro rozdělelní $\mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ vždy automaticky předpokládáme, že je regulární.

Věta A.3 *Nechť* $\boldsymbol{X} \sim \mathsf{N_p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *a nechť* $\boldsymbol{a}$ *je vektor konstant. Potom náhodná veličina* $z = \boldsymbol{a'X}$ *má normální rozdělení s parametry* $(\boldsymbol{a'\mu}, \boldsymbol{a'\Sigma a})$.

Důkaz. Viz [6]. □

Věta A.4 (zákon velkých čísel) *Nechť X_n, $n \in \mathbb{N}$ jsou nezávislé reálné náhodné veličiny s konečným rozptylem a čísla $0 < b_1 \leq b_2 \leq \ldots$, $b_n \to \infty$ jsou taková, že $\sum_{n=1}^{\infty} \frac{\mathsf{var} X_n}{b_n^2} < \infty$, potom*

$$\frac{1}{b_n} \sum_{k=1}^{n} (X_k - \mathsf{E} X_k) \to 0 \text{ skoro jistě.}$$

Důkaz. Viz [4]. □

Důsledek A.5 *Nechť $X_1, X_2, \ldots$ jsou nezávislé náhodné veličiny, které mají stejné rozdělení s konečnou střední hodnotou μ. Jestliže $n \to \infty$, pak*

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{P} \mu.$$

Literatura

[1] Anděl J.: *Základy matematické statistiky*, Matfyzpress, Praha, 2005.

[2] Hastie T., Tibshirani R., Friedman J.: *The elements of statistical learning: data mining, inference, and prediction*, Springer, New York, 2001

[3] Johnson R. A., Wichern D. W.: *Applied multivariate statistical analysis*, Prentice-Hall, Englewood Cliffs, 1982.

[4] Lachout P.: *Teorie pravděpodobnosti*(skripta), MFF UK, Praha, 1998.

[5] red. Poleno, Z. : *Lesnický naučný slovník, I. a II. díl*, Ministerstvo zemědělství, Praha, 1994.

[6] Rencher A. C.: *Multivariete statistical inference and applications*, John Wiley, New York, 1998.