

Posudek

vedoucího oponenta
diplomové bakalářské práce

Autor: **Tomáš Krupka**

Název práce: **SVM classifiers and heuristics for feature selection**

Jméno oponenta práce: **doc. RNDr. Ing. Miloš Kopa, Ph.D.**

Matematická úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Grafická, jazyková a formální úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Výsledky:

originální původní i převzaté netriviální kompilace citované z literatury opsané

Použité metody:

nestandardní standardní obojí

Aplikovatelnost:

přínos pro teorii přínos pro praxi přínos pro praxi i teorii bez přínosu nedovedu posoudit

Věcné chyby:

téměř žádné vzhledem k rozsahu a pojednávanému tématu přiměřený počet méně podstatné četné závažné

Tiskové chyby:

téměř žádné vzhledem k rozsahu a pojednávanému tématu přiměřený počet četné

Celková úroveň práce:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Připomínky a vyjádření oponenta:

Předložená práce je napsána jasně, stručně a srozumitelně i když práci lze vytknout slabší formální úroveň projevující se nepřesným značením a formulacemi, zejména:

1, Asi by bylo lepší nadrovinu na konci str. 3 označit “H(c)” protože závisí na c.

2, V definici 3: zápis $f(R^n) = L$ není jasný. Co přesně se tím myslí?

3, Když dáme dohromady definiční vztahy z Definice 3 a 4 tak dostáváme pro libovolné x :

a, $y = \text{sgn}(f(x))$ implikuje že $y = -1$ nebo $y = 1$

b, $h(x) = y$ implikuje $\text{sgn}(h(x)) = \text{sgn}(y) = y$ protože $y = \pm 1$.

c, Tedy platí $\text{sgn}(h(x)) = y$ a podle definice 4 tedy každý vstupní vektor (x, y) je korektně klasifikován což nedává smysl.

Zdá se tedy, že y má v dvou po sobě jdoucích definicích odlišný význam, což je dosti zavádějící a není jasné, co tedy přesně v dalším textu y znamená.

4, V (1.4) chybí ve druhém řádku znaménko mínus u prostředního výrazu. Podobně ve vzorci (1.10). Naštěstí to nemá vliv na další odvozování.

5, Str. 9, ř. 9: Věta “Not every function.... is a kernel.” je zavádějící. Žádná taková funkce nemůže být jádrem. Ale některé jej můžou indukovat (generovat).

6, Str. 13, 1. krok algoritmu: Nemělo by tam spíše být $\delta > 0$ místo $\eta > 0$?

7, Str. 20: Co přesně znamená “better solution” ve větě: “If a better solution is found, it is used...”

8, Str. 26, tabulka 3.1: Co znamenají podbarvené hodnoty? Proč u “arrive” není podbarvena hodnota 82.6? Proč u “throw” není podbarvena žádná hodnota?

Dále si myslím, že pokud algoritmus používá cross vzorek nebo validační vzorek na učení (testuje tam které proměnné vyhodit), je potřeba poté finální performance validovat na dalším odděleném vzorku. Jinak to není fér a výsledky můžou být zavádějící.

Celkově ale předloženou práci považuji za velmi dobrou a náročností nadstandardní, a proto ji navrhuji uznat jako práci bakalářskou.

V Praze, 25. 8. 2016

doc. RNDr. Ing. Miloš Kopa, Ph.D.