

In machine learning applications with a large number of computer-generated features, a selection of just a subset of features is often desirable. The Recursive Feature Elimination (SVM-RFE) algorithm proposed by Guyon et al. (2002) employs the mechanism of selecting the features based on their contribution to an SVM model decision rule, and has proven a state-of-the-art performance on the Gene Selection for Cancer Classification task (Tan et al. (2010)). This thesis expands on that work, and proposes a novel modification of the SVM-RFE feature selection method called Evaluation-Based RFE (EB-RFE). This heuristic significantly improves the performance of the SVM classifier in comparison to the original SVM-RFE on the studied machine learning task. In addition to the performance gain, the proposed algorithm has also, in experimental use, proven to have two other desirable properties. Firstly, EB-RFE produces much smaller feature subsets than SVM-RFE, which leads to more compact models. Secondly, unlike SVM-RFE, the EB-RFE heuristic is easily scalable with the computational time well beyond the possibilities of current high-end consumer CPUs.