FEATURE SELECTION USING CPL CRITERION FUNCTIONS

Tomasz Łukaszuk

Faculty of Computer Science, Bialystok Technical University

The feature selection is the technique, commonly used in machine learning, of selecting a subset of the most important features for building robust learning models. By removing most irrelevant and redundant features from the data, feature selection helps improve the performance of models constructed on the base of that data. In other words the feature selection means neglecting such measurements (features) which have no significant influence on the final decisions [**?**].

Dimensionality reduction is a preprocessing step commonly applied in pattern recognition and classification applications. It makes easier the next data analysis steps by alleviating the effect of the curse of dimensionality, enhancing generalization capability, speeding up learning process and improving model interpretability. Feature selection also helps people to acquire better understanding about their data by telling them which features are the important features.

This work is engaged in the feature selection by minimization of a special convex and piece-wise linear (CPL) criterion function. The minimization process allows to calculate the parameters of hyperplane saparated the lerning sets and to find the best set of features ensured the linear separability of them at once. Moreover the goal of the work is to make a comparison of described method expermiental results with the NIPS2003 Feature Selection Challenge participant's methods results.

NIPS is the acronym of Neural Information Processing Systems. It is the annual conference name taken place in Vancouver, Canada from 1987. Its topics span a wide range of subjects including neuroscience, learning algorithms and theory, bioinformatics, image processing, and data mining [**?**].

In 2003 within the framework of NIPS Conference took place the challenge in feature selection. The organizers provided participants with five datasets from different application domains and called for classification results using a minimal number of features. All datasets are two-class classification problems. The data were split into three subsets: a training set, a validation set, and a test set. The participants could submit prediction results on the validation set and get their performance results and ranking on-line for a period of 12 weeks. On December 1st, 2003, the participants had to turn in their results on the test set. The validation set labels were released at that time. On December 8th, 2003, the participants could make submissions of test set predictions, after having trained on both the training and the validation set [**?**]. Performance was assessed using several metrics, such as:

- the balanced error rate (the average of the error rate of the positive class and the error rate of the negative class),
- area under the ROC curve (he ROC curve is obtained by varying a threshold on the discriminant values (outputs) of the classifier, The curve represents the fraction of true positive as a function of the fraction of false negative),
- fraction of features selected,
- fraction of probes (random artificially generated features) found in the feature set selected.

[1] The work was partially financed by the KBN grant 3T11F01130, and by the grant S/WI/2/08 from the Białystok Technical University.

The NIPS 2003 challenge on feature selection is over, but the website of the challenge is still open for post-challenge submissions. One can compare results by his own method with the challenge participant's methods results.

References

[1] Bobrowski L.: Design of Piecewise Linear Classifiers from Formal Neurons by Some Basis Exchange Technique, pp. 863–870 in: Pattern Recognition, 24(9), 1991.

[2] Bobrowski L., Łukaszuk T.: Selection of the linearly separable feature subsets, pp.544-549 in: Artificial intelligence and soft computing: ICAISC'2004, eds. Leszek Rutkowski, Jörg Siekmann, Ryszard Tadusiewicz, Lotfi A. Zadeh, Lecture Notes in Computer Science, vol.3070, 2004.

[3] Bobrowski L.: Eksploracja danych oparta na wypukłych i odcinkowo-liniowych funkcjach kryterialnych, Wyd. Politechniki Białostockiej, Białystok, 2005.

[4] Duda, O.R., Heart, P.E., Stork D.G.: Pattern Classification, Second edition, John Wiley & Sons, 2001.

[5] `http://nips.cc/`

[6] `http://www.nipsfsc.ecs.soton.ac.uk/`