

Machine Learning Exercise Lecture 4

Christian Hilligsøe, Laura Nyrup and Trine Jensen, grp. 234

27-09-2018

Solution Summary

In this exercise we have performed classification on the entire MNIST dataset, by use of the Support Vector Machine (SVM) method. The Classification is performed in python using the `sklearn` library.

In order to reduce the complexity of the calculation, sklearn's Linear Discriminant Analysis (LDA) is used to obtain a dimensionally reduced dataset of $dim = n \times 9$.

In the script a linear kernel and a Radial basis function (C=5, gamma=0.05) (RBF) is used as kernels.

SVM.SVC implement the "one-against-one" approach for multi- class classification. If `n_class` is the number of classes, then `n_class · (n_class - 1)/2` classifiers are constructed and each one trains data from two classes.¹

Results

LDA with a dimensional reduction to 9

Kernel: "linear"

Accuracy for:

All classes is 89.33%

Class 0 is 96.43%

Class 1 is 96.48%

Class 2 is 87.60%

Class 3 is 86.83%

Class 4 is 92.67%

Class 5 is 83.30%

Class 6 is 93.32%

Class 7 is 88.33%

Class 8 is 81.72%

Class 9 is 85.33%

LDA with a dimensional reduction to 9

kernel: "RBF"

Accuracy for: All classes is 92.54%

Class 0 is 98.06%

Class 1 is 97.97%

¹<http://scikit-learn.org/stable/modules/svm.html#svm-classification>

Class 2 is 91.96%
Class 3 is 90.30%
Class 4 is 93.79%
Class 5 is 88.57%
Class 6 is 94.78%
Class 7 is 92.32%
Class 8 is 87.58%
Class 9 is 89.10%

Further analysis were done with computation on data reduced to lower dimensions, but this only leads to a worse accuracy, still with RBF being the best kernel.