Machine Learning - Assignment 6

Christoffer Thrysøe - dfv107

January 8, 2017

1 Kernel-induced metric

First we note that:

$$\begin{aligned} ||\phi(x) - \phi(z)|| &= \sqrt{\langle \phi(x) - \phi(z), \phi(x) - \phi(y) \rangle} \\ &= \sqrt{\langle k(x,\cdot) - k(z,\cdot), k(x,\cdot) - k(z,\cdot) \rangle} \end{aligned}$$

Applying rules of the inner product on the inner term above yields the following:

$$\begin{aligned} \langle k(x,\cdot) - k(z,\cdot), k(x,\cdot) - k(z,\cdot) \rangle &= \langle k(x,\cdot), k(x,\cdot) - k(z,\cdot) \rangle - \langle k(z,\cdot), k(x,\cdot) - k(z,\cdot) \rangle \\ &= \langle k(x,\cdot) - k(z,\cdot), k(x,\cdot) \rangle - \langle k(x,\cdot) - k(z,\cdot), k(x,\cdot) \rangle \\ &= \langle k(x,\cdot) - k(x,\cdot) \rangle - \langle k(z,\cdot), k(x,\cdot) \rangle - \langle k(x,\cdot), k(z,\cdot) \rangle + \langle k(z,\cdot), k(z,\cdot) \rangle \\ &= \langle k(x,\cdot), k(x,\cdot) \rangle - 2\langle k(z,\cdot), k(x,\cdot) \rangle + \langle k(z,\cdot), k(z,\cdot) \rangle \\ &= k(x,x) - 2k(z,x) + k(z,z) \end{aligned}$$

Inserting this into the original equation we get the following:

$$||\phi(x) - \phi(z)|| = \sqrt{k(x,x) - 2k(z,x) + k(z,z)}$$

Thus concluding the proof

2. SVM in practice

2.1 Data normalization

The data normalizing procedure is located in the file `normalize.py`. The code has been reused from assignment 2.

Table 1: Mean and variance of the training and test data

	Training Data	Transformed Test Data
Mean	21.823	0.126
Variance	3316.841	1.572

2.2 Model selection using grid-search

The implementation for cross-validation is found in the file `svmc.py` I have used the `Libsvm` implementation of support vector machine, using the radial basis function / Gaussian as a kernel. I have evaluated the following values for C:

$$[0.01, 0.1, 1, 10, 100, 1000, 10000, 100000, 1000000, 10000000]$$

and for γ:

$$[1.0 \cdot 10^{-6}, 1.0 \cdot 10^{-5}, 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000]$$

For the cross validation, the training data is split into 5 roughly equal sized arrays. Each partition is evaluated as the validation set at turn. The error is then measured for all combinations of C and γ. The combination of C and γ, resulting in the lowest error, is then stored. After running the cross validation for all validation sets, the combination of parameters resulting in the lowest validation error over all validation sets is then returned. This combination of parameters is then used to evaluate the generalization error. The following table shows the result of running the cross validation algorithm:

Table 2: Best hyperparameter configuration

Avg. Error	γ	C
0.11158	0.1	1

The hyperparameters were chosen by identifying the smallest average error of all parameter combinations across 5-cross-validation. This is identified by accumulating the grid matrices, and taking the average over five folds, the index of the entry with the smallest average error, identifies the best hyper parameters.
I ran the SVM using the above hyperparameters, using both raw data and normalized data for training and testing, the result was:

Table 3: Generalization error

	Error
Test Data	0.216
Normalized test Data	0.134

Thus normalizing the data greatly improves the generalization.