STAD68 Advanced Machine Learning and Data Mining
Fall 2016

Homework 2

Due: 20 October 2016

Homework submission: Homework should be submitted electronically via Blackboard by 11.59pm on the due date.

Allowed resources: You may use the course notes, the books listed on the syllabus, and Wikipedia. Do not Google for the answer.

Problem 2 (Combining kernels)

It was already mentioned in class that kernel functions can be combined and modified in various ways to obtain new kernel functions. In this problem, we will convince ourselves that this is indeed true in two simple cases. Recall that a function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a kernel if there is *some* function $\phi : \mathbb{R}^d \to \mathcal{F}$, into *some* space $\mathcal{F}$ with scalar product $\langle\, . \,, \,. \,\rangle_{\mathcal{F}}$, such that

$$k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{F}}$$

for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$.

Homework problems:
Let $k_1(\mathbf{x}, \mathbf{x}')$ and $k_2(\mathbf{x}, \mathbf{x}')$ be kernels on $\mathbb{R}^d$. For all problems below, you can assume $\mathcal{F} = \mathbb{R}^D$ for some $D > d$.

1. Show that, for any positive real number a, $k(\mathbf{x}, \mathbf{x}') := ak_1(\mathbf{x}, \mathbf{x}')$ is a kernel.
2. Show that $k(\mathbf{x}, \mathbf{x}') := k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}')$ is a kernel.
3. Show that, for any positive integer p, $k(\mathbf{x}, \mathbf{x}') := k_1(\mathbf{x}, \mathbf{x}')^p$ is a kernel.

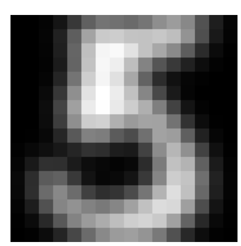
Problem 2 (SVM)

In this problem, we will apply a support vector machine to classify hand-written digits. You do not have to implement the SVM algorithm: The R library e1071 provides an implementation, see

`http://cran.r-project.org/web/packages/e1071/index.html`

You may also use an SVM implementation in python (e.g., in scikit-learn).

Download the digit data set from Blackboard. The zip archive contains two files: Both files are text files. The file `uspsdata.txt` contains a matrix with one data point (= vector of length 256) per row. The 256-vector in each row represents a 16×16 image of a handwritten number. The file `uspscl.txt` contains the corresponding class labels. The data contains two classes—the digits 5 and 6—so the class labels are stored as -1 and +1, respectively. The image on the right shows the first row, re-arranged as a 16×16 matrix and plotted as a gray scale image.



Homework problems:

1. Reproduce the plot of the first digit, but replace the upper right pixel with a gray pixel.

2. Randomly select about 20% of the data and set it aside as a test set. Train a linear SVM with soft margin. Cross-validate the margin parameter using 5-fold cross validation. Plot the cross-validation estimates of the misclassification rate. Plot the rate as a function of the margin parameter. Report the test set estimates of the misclassification rate, with the parameter values you have selected by cross validation.

3. Train an SVM with soft margin and RBF kernel. You will have to cross-validate both the soft-margin parameter and the kernel bandwidth. Plot the cross-validation estimates of the misclassification rate. Plot the rate as a function of the margin parameter and bandwidth. Report the test set estimates of the misclassification rate, with the parameter values you have selected by cross validation.

4. Compare the two classifiers. Is a linear SVM a good choice for this data, or should we use a non-linear one?

5. (Bonus) Optimize the margin parameter and kernel bandwidth on the training data. (That is, don't do cross validation.) Compare your classification performance with the performance above. Do you suffer from overfitting?

6. (Extra bonus) Implement the SVM yourself using a package for mathematical optimization for your language of choice.