Practical Session 7

Soft-Margin SVM and Kernels

1 Soft-Margin SVM

Problem 1: What is the connection between soft-margin SVM and logistic regression?

2 Gaussian kernel

Problem 2: One of the nice things about kernels is that new kernels can be constructed out of already given ones. Use the five kernel construction rules from the lecture to prove that the function

$$K(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{y}|^2}{2\sigma^2}\right)$$

is a kernel.

(Hint: Use the Taylor expansion of the exponential function to prove that $\exp(K_1(\boldsymbol{x}, \boldsymbol{y}))$ is a kernel if $K_1(\boldsymbol{x}, \boldsymbol{y})$ is a kernel.)

2nd hint: It might help to apply the rule $K_3(\phi(\boldsymbol{x}), \phi(\boldsymbol{y}))$ with the linear kernel $K_3(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T \boldsymbol{y}$ and consider a feature map $\phi(z)$ with only one feature.

3 Stacking feature maps

Suppose you have found a feature map $\boldsymbol{\theta} : \mathbb{R}^n \to \mathbb{R}^m$ that transforms your data into a feature space in which a SVM with a Gaussian kernel works well. However computing the feature map $\boldsymbol{\theta}(\boldsymbol{x})$ is computationally expensive and luckily you discover an efficient method to compute the scalar product $K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\theta}(\boldsymbol{x})^T \boldsymbol{\theta}(\boldsymbol{y})$ in your feature space without having to compute $\boldsymbol{\theta}(\boldsymbol{x})$ and $\boldsymbol{\theta}(\boldsymbol{y})$ explicitly.

Problem 3: Show how you can use the scalar product $K(\boldsymbol{x}, \boldsymbol{y})$ to efficiently compute the Gaussian kernel in your feature space, that is

$$K_g(\boldsymbol{\theta}(\boldsymbol{x}), \boldsymbol{\theta}(\boldsymbol{y}))$$

where

$$K_g(\boldsymbol{a}, \boldsymbol{b}) = \exp\left(-\frac{|\boldsymbol{a} - \boldsymbol{b}|^2}{2\sigma^2}\right)$$

is the Gaussian kernel.

4 Unsuitable Kernels

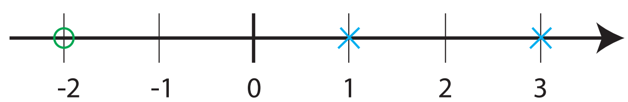
Consider a SVM without slack variables using the kernel $K(x, y) = |x||y|$ where $x, y \in \mathbb{R}$.

Problem 4: Show that the function $K(x, y)$ is indeed a kernel and write down a feature space $\phi(x)$ corresponding to this kernel.

Problem 5: Write down the set of classification functions that can be learned by an SVM using the kernel $K(x, y)$. Make sure that the set contains each classification function only once.

Consider the following data set. The ith data point is given by x_i and the corresponding class is $y_i \in \{-1, +1\}$.

i	x_i	y_i
1	-2	-1
2	1	$+1$
3	3	$+1$



Problem 6: What would happen if you try to solve the dual problem for fitting an SVM without slack variables to this data set? Explain your answer.

5 Kernelized k-nearest neighbors

To classify the point $\boldsymbol{x}$ the k-nearest neighbors finds the k training samples $\mathcal{N} = \{\boldsymbol{x}^{(s_1)}, \boldsymbol{x}^{(s_2)}, \dots, \boldsymbol{x}^{(s_k)}\}$ that have the shortest distance $||\boldsymbol{x} - \boldsymbol{x}^{(s_i)}||_2$ to $\boldsymbol{x}$. Then the label that is mostly represented in the neighbor set $\mathcal{N}$ is assigned to $\boldsymbol{x}$.

Problem 7: Formulate the k-nearest neighbors algorithm in feature space by introducing the feature map $\boldsymbol{\phi}(\boldsymbol{x})$. Then rewrite the k-nearest neighbors algorithm so that it only depends on the scalar product in feature space $K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{\phi}(\boldsymbol{y})$.