Assignment 2: Support Vector Machines
Machine Learning

Deadline: Friday 3 Nov 2017, 21:00

Introduction

In this assignment, you will further deepen your understanding of Support Vector Machines (SVMs). Please provide a latex based report in the PDF format.

Your report must be archived in a file named firstname.lastname and uploaded to the iCorsi website before the deadline expires. Late submissions will result in 0 points.

Where to get help

We encourage you to use the tutorials to ask questions or to discuss exercises with other students. However, do not look at any report written by others or share your report with others. Violation of that rule will result in 0 points for all students involved. You can contact your TAs using the emails *imanol@idsia.ch* and *boris@idsia.ch*.

Grading

The assignment consists of eight tasks totalling at 100 points. Bonus points are not necessary to achieve the maximal grade. You will be awarded 5 bonus points for a well-presented report (clear writing, annotated equations).

Tasks

We consider an input vector space $\mathbf{x} \in \mathbb{R}^N$ and a classifier $sign(f(\mathbf{x}))$ assigning a label of -1 or $+1$ to $\mathbf{x}$ based on a linear function $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$, where $\mathbf{w}^T\mathbf{x}$ denotes the inner product of vectors $\mathbf{w}$ and $\mathbf{x}$.

1. **(10 points)** Explain in your own words how an SVM roughly works. Make sure to add key characteristics, advantages, and disadvantages.

2. **(5 points)** What is a hyperplane? What are Support Vectors? Give a few examples.

3. **(15 points)** Explain the objective functions of hard-margin and soft-margin support vector machine training as well as the constraints of the corresponding optimization problems. If you like, refer to a sketch.

4. **(5 points)** Give an overview of the hyper-parameters of an SVM and what their effect is.

5. **(10 points)** The simple classifier function of a support vector machine is linear. Explain how we can still use it to separate non-linearly separable data.

6. **(10 points)** Explain the relation between a kernel function and a feature map. What are those? What does the kernel function compute? What is known as the "kernel trick"?

7. **(15 points)** Explain the free parameter of the Gaussian Kernel and what its effect is. Give a simple example of a complex non-linear separable problem and explain how it could be solved using an SVM with Gaussian Kernels.

8. **(30 points)** Consider the cubic kernel

$$k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}; \quad k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y})^3 = (x_1 y_1 + x_2 y_2)^3 \tag{1}$$

What is the corresponding feature map that belongs to that kernel? That is what is the function

$$\phi : \mathbb{R}^2 \to \mathbb{R}^k. \tag{2}$$

Show now that the following relation holds:

$$k(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y}) \tag{3}$$