Week 8 Homework Submission
SVM and Kernels

Iuliia Skobleva
Matriculation Number 03723809

December 8, 2019

SVM

Problem 1

The main similarity between the SVM and perceptron is their goal, which is to find a hyperplane, which will separate the data. The differences are the following: the perceptron algorithm terminates once the hyperplane, which would linearly separate the points, is found. The might be infinitely many of these "optimal" hyperplanes, but the in the end, the result is the first one found. Using SVM, we are able to find a hyperplane with the widest margin, which will separate the data. There exists only one separating hyperplane in this case.

Kernels

Problem 5

We have to prove that the following kernel is valid:

$$k(\mathbf{x_1}, \mathbf{x_2}) = \sum_{i=1}^{N} a_i(\mathbf{x_1^T x_2})^i + a_0 \tag{1}$$

with $\mathbf{x_1}, \mathbf{x_2} \in \Re^d$. The Mercer's theorem states that a kernel is valid if it gives rise to a symmetric, positive semidefinite kernel matrix for any input data $\mathbf{X}$. In our case, we have two data points $\mathbf{x_1}$ and $\mathbf{x_2}$, meaning the Gram matrix is a one by one matrix. Next,

1. $k(\mathbf{x_1}, \mathbf{x_2}) = k(\mathbf{x_2}, \mathbf{x_1})$, because the scalar product is symmetric.

2. the Gram matrix is positive semidefinite, because $\forall \alpha \in \Re \; \alpha k(\mathbf{x_2}, \mathbf{x_1})\alpha \geq 0$. This holds, because the scalar product in (1), $\mathbf{x_1^T x_2}$, is non-negative, as well as its product with $a_i \geq 0$ taken to the power of i. The sum of all these terms remains non-negative and the addition of $a_0 \geq 0$ doesn't change the sign.

Therefore, the Mercel's theorem is fulfilled and $k(\mathbf{x_1}, \mathbf{x_2})$ is a valid kernel.

Problem 6

The goal is to find the feature transformation $\phi(x)$ corresponiding to the kernel:

$$k(x_1, x_2) = \frac{1}{1 - x_1 x_2} \tag{2}$$

with $x_1, x_2 \in (0, 1)$.

The hint (*consider an infinite-dimensional feature space*) and the fact that x_1, x_2 lies in range $(0, 1)$ reminds us of the geometric series, which looks like this:

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1 - r} \tag{3}$$

where $r \in (0, 1)$. Basically, we can write down the kernel as:

$$\begin{aligned} k(x_1, x_2) &= \sum_{k=0}^{\infty} x_1^k x_2^k \\ &= \frac{1}{1 - x_1 x_2} \end{aligned} \tag{4}$$

What kind of transformation can lead to this result? We can write down (4) as:

$$\begin{aligned} k(x_1, x_2) &= \sum_{k=0}^{\infty} x_1^k x_2^k \\ &= 1 + x_1 x_2 + x_1^2 x_2^2 + ... \end{aligned} \tag{5}$$

This in tern looks similar to a scalar product between two vectors, which are functions of x_1 and x_2 respectively. Indeed, if we set

$$\boldsymbol{\phi}(x) = \begin{pmatrix} 1 \\ x \\ x^2 \\ ... \end{pmatrix} \tag{6}$$

where $\boldsymbol{\phi}(x)$: $(0,1) \longrightarrow \Re^{\infty}$. The scalar product between $\boldsymbol{\phi}(x_1)$ and $\boldsymbol{\phi}(x_2)$ leads to the result in equation 5.