# Assignment 1 : CS-E4830 Kernel Methods in Machine Learning 2017

The **deadline** for this assignment is **Thursday 21.09.2017 at 4pm**. If you have **questions** about the assignment, you can ask them in the 'General discussion' section on MyCourses or attend the Q&A session on 15.09.17 at 8:30 am in U1/U154, Otakaari 1. We will have a tutorial session regarding the **solutions** of this assignment on 22.09.17 at 8:30 am in U1/U154. The solutions will also be available in MyCourses.

The report for the assignment should contain your proposed solution for the pen and paper exercises. Regarding the computer exercise, you should explain what you have done and include figure(s) (correctly annotated with legend and axe titles) or table(s) that summarize your results. You should also comment about the results you obtained. The report **and** the code (in Matlab, Python or R) must be returned as a single .zip file to MyCourses (naming convention: lastname-firstname-assignment1.zip). In case you have hand-written solutions for the pen and paper exercises you can scan them and add them in your report or submit them to any of the course teaching assistants (Parisa, Celine or Sandor) in their office (A319 or A323).

## Pen and paper exercise

### Kernel Computation

Let $\kappa_1 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $\kappa_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be two kernel functions. Let $\mathbf{K}_1$ and $\mathbf{K}_2$ be the corresponding kernel matrices over some data $X$ such that $[\mathbf{K}_1]_{i,j} = \kappa_1(x_i, x_j)$ and $[\mathbf{K}_2]_{i,j} = \kappa_2(x_i, x_j)$ for $x_i, x_j \in X$. **State and prove** whether the following statements are **true or false**. Please support your answer with a mathematical proof or a counter example.

**Question 1** (1 point):

If $\mathbf{K}$ is a positive semidefinite kernel matrix, then all its entries must be positive.

**Question 2** (1 point):

If $\mathbf{K} = a\mathbf{K}_1 + b\mathbf{K}_2$, where $a, b \in \mathbb{R}^+$, then $\mathbf{K}$ is a valid kernel matrix.

**Question 3** (1 point):

If $\mathbf{K} = \mathbf{K}_1 - \mathbf{K}_2$, then $\mathbf{K}$ is a valid kernel matrix.

### Parzen window classifier

In this exercise we consider a binary classification problem. Let $S = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$, $y_1, \ldots, y_m \in \{-1, +1\}$ be the training set that contains $m_+$ positive examples and $m_-$

negative examples. Let $I = \{1, \ldots, m = m_+ + m_-\}$ be the indices of the training examples. We note $I^+ = \{i \in I | y_i = +1\}$ the set containing the indices of the positive training examples. We define similarly $I^- = \{i \in I | y_i = -1\}$ for the negative training examples.

$\kappa$ is a kernel defined on $\mathcal{X} \times \mathcal{X}$, and $\phi$ is a feature map associated with this kernel.

Let $\mathbf{c}_+ = \frac{1}{m_+} \sum_{i \in I^+} \phi(\mathbf{x}_i)$ and $\mathbf{c}_- = \frac{1}{m_-} \sum_{i \in I^-} \phi(\mathbf{x}_i)$ be the means of the two classes in the feature space. Given a new point $\mathbf{x} \in \mathcal{X}$ to classify, the idea of the Parzen window classifier is to assign $\mathbf{x}$ to the closest class in the feature space:

$$h(\mathbf{x}) = \begin{cases} +1 \text{ if } ||\phi(\mathbf{x}) - \mathbf{c}_-||^2 > ||\phi(\mathbf{x}) - \mathbf{c}_+||^2 \\ -1 \text{ otherwise.} \end{cases}$$

The function $h$ can be expressed using the sign function:

$$h(\mathbf{x}) = \operatorname{sgn}(||\phi(\mathbf{x}) - \mathbf{c}_-||^2 - ||\phi(\mathbf{x}) - \mathbf{c}_+||^2).$$

**Question 4** (2 points):

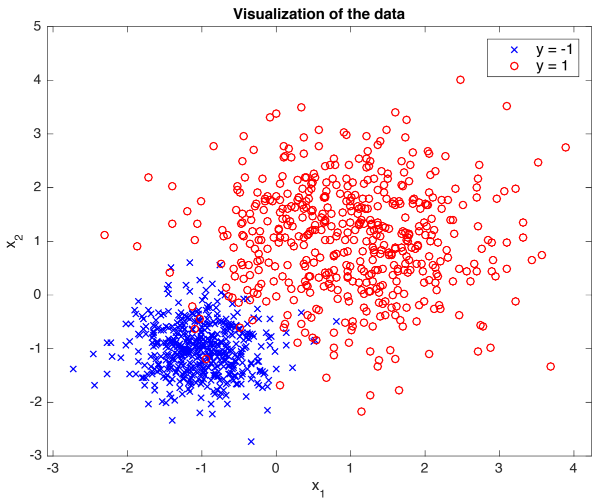
Show that the function $h$ can be written as: $h(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^m \alpha_i \kappa(\mathbf{x}, \mathbf{x}_i) + b\right)$, where

$$b = \frac{1}{2m_-^2} \sum_{i,j \in I^-} \kappa(\mathbf{x}_i, \mathbf{x}_j) - \frac{1}{2m_+^2} \sum_{i,j \in I^+} \kappa(\mathbf{x}_i, \mathbf{x}_j),$$

$$\alpha_i = \begin{cases} \dfrac{1}{m_+} & \text{if } y_i = +1 \\ \dfrac{-1}{m_-} & \text{if } y_i = -1 \end{cases}$$

(Hint: use the kernel trick $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle = \kappa(\mathbf{x}_i, \mathbf{x}_j)$.)

## Computer exercise




In this exercise, the Parzen window classifier will be applied on a synthetic classification dataset. In this dataset, each data point is represented by a vector of length 2. The training set contains 800 examples, while the test set contains 200 examples. The file 'data_all.mat' contains the inputs and binary outputs for the training set (X_train and y_train), as well as the inputs and outputs for the test set (X_test, y_test). If you are not using Matlab, you can load these files separately: 'X_train.txt', 'X_test.txt', 'y_train.txt' and 'y_test.txt'.

**Question 5** (1.5 points):

Write a function that computes a sub-matrix of the kernel matrix $\mathbf{K}$ of a Gaussian kernel function $\kappa$ between a matrix $\mathbf{X}$ of size $n \times d$ and a matrix $\mathbf{Z}$ of size $m \times d$. This sub-matrix should be a matrix of size $n \times m$ defined as:

$$\begin{bmatrix} \kappa(\mathbf{x}_1, \mathbf{z}_1) & \kappa(\mathbf{x}_1, \mathbf{z}_2) & \dots & \kappa(\mathbf{x}_1, \mathbf{z}_m) \\ \kappa(\mathbf{x}_2, \mathbf{z}_1) & \kappa(\mathbf{x}_2, \mathbf{z}_2) & \dots & \kappa(\mathbf{x}_2, \mathbf{z}_m) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(\mathbf{x}_n, \mathbf{z}_1) & \kappa(\mathbf{x}_n, \mathbf{z}_2) & \dots & \kappa(\mathbf{x}_n, \mathbf{z}_m) \end{bmatrix}.$$

The function should take in input:

- a matrix $\mathbf{X}$ of size $n \times d$,
- a matrix $\mathbf{Z}$ of size $m \times d$,
- the parameter $\sigma$ of the Gaussian kernel.

**Question 6** (2 points):

Write a function that computes the predicted values of the function $h$ defined in Question 4 on some test set. This function should take in input:

- the kernel matrix of the Gaussian kernel between the matrices **X**_train and **X**_train,
- the kernel matrix of the Gaussian kernel between the matrices **X**_train and **X**_test,
- the vector containing the training labels (y_train).

The kernel matrix between the training input matrices is needed for computing $b$ (see Question 4) and the kernel matrix between the training and test input matrices is needed for computing the predictions (h($\mathbf{x}$)).

**Question 7** (1.5 points):

Experiment the Parzen window classifier using a Gaussian kernel. In order to evaluate the performances, compute the accuracy (percentage of test examples that are correctly classified) obtained on the training and test set for different choices of the kernel parameter $\sigma$ ($\sigma$ = [0.01, 0.05, 0.1, 0.5, 1, 5, 10]).

Plot the training and test accuracies obtained for different values of the kernel parameter $\sigma$. Describe the behavior of the accuracy when the parameter of the Gaussian kernel varies. Write your comments in your report and include the figure(s) you obtained.

Hint: The test accuracy should be around 95% for the best kernel parameter.