Exercises for the course
Machine Learning 2
Summer semester 2024

Fachgebiet Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 6

Exercise 1: Convolution Kernel (20 P)

Let x, x' be two univariate real-valued discrete signals, that we consider in the following to be infinite-dimensional. We consider a discrete convolution over these two signals

$$[x * x']_t = \sum_{\tau=-\infty}^{\infty} x(\tau) \cdot x'(t-\tau)$$

which also produces an infinite-dimensional output signal. We then define the 'convolution kernel' as:

$$k(x, x') = \|x * x'\|^2 = \sum_{t=-\infty}^{\infty} ([x * x']_t)^2.$$

(a) *Show* that the convolution kernel is positive semi-definite, that is, show that

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k(x_i, x_j) \geq 0$$

for all inputs $x_1, \ldots, x_N$ and choice of real numbers $c_1, \ldots, c_N$.

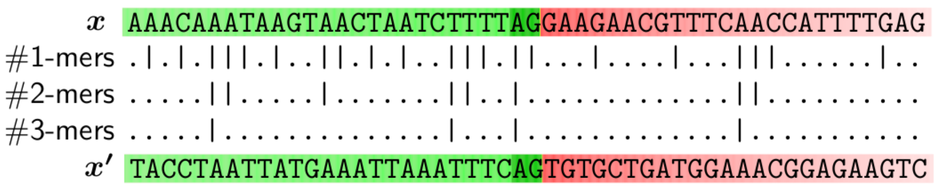
(b) *Give* an explicit feature map for this kernel.

Exercise 2: Weighted Degree Kernels (20 P)

The weighted degree kernel has been proposed to represent DNA sequences ($\mathcal{A} = \{\mathrm{G}, \mathrm{A}, \mathrm{T}, \mathrm{C}\}$), and is defined for pairs of sequences of length L as:

$$k(x, z) = \sum_{m=1}^{M} \beta_m \sum_{l=1}^{L+1-m} I(u_{l,m}(x) = u_{l,m}(z)).$$

where $\beta_1, \ldots, \beta_M \geq 0$ are weighting coefficients, and where $u_{l,m}(x)$ is a substring of x which starts at position l and of length m. The function $I(.)$ is an indicator function which returns 1 if the input argument is true and 0 otherwise.



(a) *Show* that k is a positive semi-definite kernel. That is, show that

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k(x_i, x_j) \geq 0$$

for all inputs $x_1, \ldots, x_N$ and choice of real numbers $c_1, \ldots, c_N$.

(b) *Give* a feature map associated to this kernel for the special case $M = 1$.

(c) *Give* a feature map associated to this kernel for the special case $M = 2$ with $\beta_1 = 0$ and $\beta_2 = 1$.

Exercise 3: Fisher Kernel (20 P)

The Fisher kernel is a structured kernel induced by a probability model $p_\theta(x)$. While it is mainly used to extract a feature map of fixed dimensions from structured data on which a structured probability model readily exists (e.g. a hidden Markov model), the Fisher kernel can in principle also be derived for simpler distributions such as the multivariate Gaussian distribution.

The probability density function of the Gaussian distribution in $\mathbb{R}^d$ of mean μ and covariance Σ is given by:

$$p(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\Big(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\Big)$$

In this exercise, we consider the covariance matrix Σ to be fixed, and therefore, the only effective parameter of the model (on which the Fisher kernel is based) is the mean μ.

(a) *Show* that the Fisher kernel associated to this probability model is given by:

$$k(x, x') = (x-\mu)^\top \Sigma^{-1}(x'-\mu)$$

(b) *Give* a feature map associated to this kernel.

Exercise 4: Programming (40 P)

Download the programming files on ISIS and follow the instructions.