Medical Data Science, SS 2019
Prof. Dr. Nico Pfeifer
Chair for Methods in Medical Informatics
University of Tuebingen

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

2018-05-09

Assignment 2

Deadline: Thursday, May 23, 9:59 p.m.

This problem set is worth 32 points. You can submit in groups of two people or alone. Submit your solutions digitally by uploading to the ILIAS webpage (none of the other students can see the files you upload). Just upload a zipped folder containing all necessary files and name the folder by your last name(s). The folder should be named according to the following scheme:

`[MDS][Assignment 2]_lastname`
or
`[MDS][Assignment 2]_lastname1_lastname2`

Problem 1 (T, 15 Points)

Kernel methods.

(a) (2P) Describe shortly the concept of *alternative splicing* using the terms *exon*, *intron*, *donor* and *acceptor splice site*.

(b) (3P) Compare the *spectrum kernel* with the *weighted degree kernel with shifts*. In what sense are they similar to each other, where are the differences? Also consider the parameters that have to be specified for the kernel calculation.

(c) (3P) The oligo function is a sum of gaussians. In this context, what do the parameters of the gaussians represent? How is the similarity between two sequences determined?

Domain adaptation and multitask learning

(a) (3P) Why does the task of *domain adaptation* emerge in machine learning? Describe two approaches to domain adaptation.

(b) (3P) Explain the *multitask learning* approach of lecture three. How does the topology of the tree that corresponds to the relationships among the tasks influence the optimization?

(c) (1P) Explain the role of the Major Histocompatibility Complex (MHC) in the immune system.

Problem 2 (T, 8 Points)

Kernel functions use an implicit mapping of the data points to a potentially high-dimensional Hilbert space $\phi : x_i \rightarrow \phi(x_i)$, meaning that $k(x_i, x_j) = \left\langle \phi(x_i), \phi(x_j) \right\rangle$. Show, how we can calculate the squared euclidean distance between two samples in this Hilbert space $\|\phi(x_i) - \phi(x_j)\|^2$ without using the mapping function ϕ.

Problem 3 (P, 9 Points)

Implement the weighted degree kernel (without shifts) in Matlab, R, or python and compute and visualize kernel matrices. The data can be found here.

(a) (7P) Implement the weighted degree kernel as a function in Matlab, R, or python that takes two sequences as well as the parameter d, and the β parameters. The output should be the kernel value as defined in slide 37.

(b) (2P) Visualize the kernel matrix for $d = 3$ and $\beta_k = 2(d-k+1)/(d(d+1))$ for the 8 sequences from here. The visualization should show a 8 x 8 matrix representing the kernel values in a heat map (e.g., with imagesc() in Matlab).