Exercises for the course
Machine Learning 2
Summer semester 2016

Abteilung 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

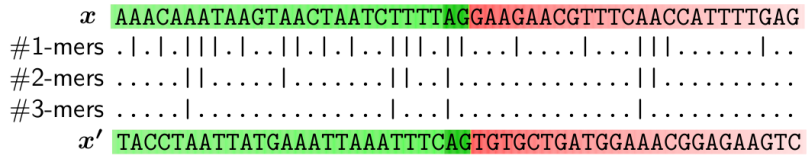
Exercise Sheet 7

Exercise 1: Weighted Degree Kernels (20 P)

We would like to implement a classifier for genes sequences (a sequence of symbols $\{A, C, T, G\}$). The weighted degree kernel is proposed for such task and is defined as:

$$k(x, x') = \sum_{m=1}^{M} \beta_m \sum_{n=1}^{N-m+1} I(u_{m,n}(x) = u_{m,n}(x')).$$

where $u_{m,n}(x)$ is a string of length m which starts at position n in sequence x, and $\beta_m \geq 0$. The symbol $I(.)$ denotes the indicator function which returns 1 if the input argument is true and 0 otherwise.



(a) *Explain* what kind of structural knowledge the weighted degree kernel seeks to incorporate.

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

$$\sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_i \alpha_j k(x_i, x_j) \geq 0$$

for all inputs $x_1, \ldots, x_K$ and choice of real numbers $\alpha_1, \ldots, \alpha_K$.

Exercise 2: Implementing Kernels for Genes Sequences and Text (80 P)

Download the IPython notebook on ISIS and follow the instructions.