

Machine Learning 2 – Group ESHG  
Assignment 07

Willi Gierke, Arik Elimelech, Mehmed Halilovic, Leon Sixt

June 9, 2017

### Exercise 1: Weighted Degree Kernels (30 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  $I(\cdot)$  denotes the indicator function which returns 1 if the input argument is true and 0 otherwise.

[illegible]

- (a) 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  $x_1, \dots, x_K$  and choice of real numbers  $\alpha_1, \dots, \alpha_K$ .

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

### Exercise 2: Programming (70 P)

Download the programming files on ISIS and follow the instructions.