

Exercise Sheet 5

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, \dots, x_N and choice of real numbers c_1, \dots, 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 ($\mathcal{A} = \{\text{G}, \text{A}, \text{T}, \text{C}\}$), 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 $\beta_1, \dots, \beta_M \geq 0$ 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 $I(\cdot)$ is an indicator function which returns 1 if the input argument is true and 0 otherwise.

x AAACAAATAAGTAACTAATCTTTTAGGAAGAACGTTTCAACCATT TTGAG
#1-mers .|. |.| |.| |.| |.| |.| |.| |.| |.| |.| |.| |..
#2-mers||.....| |.| |.....| |.....
#3-mers|.....|...|.....|.....
 x' TACCTAATTATGAAATTAAATTTTCAGTG TGCTGATGGAAACGGAGAAGTC

- (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, \dots, x_N and choice of real numbers c_1, \dots, 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:

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

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.

1 Convolutional Kernel

$$\begin{aligned}
 a) \quad & \sum_{i,j} c_i c_j h(x_i, x_j) \geq 0 \\
 &= \sum_{i,j} c_i c_j \int_{-\infty}^{\infty} (x_i * x_j)_t (x_j * x_i)_t \\
 &= \sum_{i,j} c_i c_j \sum_{t=-\infty}^{\infty} \left(\sum_{s=-\infty}^{\infty} x_i(s) \cdot x_j(t-s) \right) \left(\sum_{s=-\infty}^{\infty} x_j(s) \cdot x_i(t-s) \right) \\
 &= \sum_{t=-\infty}^{\infty} \left(\sum_{i=1}^N c_i \sum_{s=-\infty}^{\infty} x_i(s) x_j(s+t) \right) \left(\sum_{j=1}^N c_j \sum_{s=-\infty}^{\infty} x_j(s) x_i(s+t) \right) \\
 &= \sum_{t=-\infty}^{\infty} \left(\sum_{i=1}^N c_i \sum_{s=-\infty}^{\infty} x_i(s) x_j(s+t) \right)^2 \geq 0
 \end{aligned}$$

$$\begin{aligned}
 b) \quad & \langle x * x' \rangle = \sum_s \sum_t x(t) x(s+t) \sum_i x'(t) x'(s+t) \\
 &= \underbrace{\langle x * x \rangle}_{\phi(x)} \cdot \underbrace{\langle x' * x' \rangle}_{\phi(x')}
 \end{aligned}$$

2 Weighted Degree Kernels

$$\begin{aligned}
 a) \quad & \sum_{i,j} c_i c_j \sum_m \beta_m \sum_l \mathbb{1}(u_{l,m}(x) = u_{l,m}(z)) \geq 0 \\
 \Leftrightarrow & \sum_{i,j} c_i c_j \sum_{m,l} \beta_m \sum_{S \in A^m} \mathbb{1}(u_{l,m}(x) = S) \mathbb{1}(u_{l,m}(z) = S) \\
 \Leftrightarrow & \sum_{l,m,S} \beta_m \left(\sum_i c_i \mathbb{1}(u_{l,m}(x) = S) \right) \left(\sum_j c_j \mathbb{1}(u_{l,m}(z) = S) \right) \\
 \Leftrightarrow & \sum_{l,m,S} \underbrace{\beta_m}_{\geq 0} \underbrace{\left(\sum_i c_i \mathbb{1}(u_{l,m}(x) = S) \right)^2}_{\geq 0} \geq 0
 \end{aligned}$$

$$b) \quad \sum_{m,l} \beta_m \sum_{S \in A^m} \mathbb{1}(u_{l,m}(x) = S) \mathbb{1}(u_{l,m}(z) = S)$$

$$\Leftrightarrow \sum_{l,S} \sqrt{\beta} \mathbb{1}(u_l(x) = S) \sqrt{\beta} \mathbb{1}(u_l(z) = S)$$

$$\Leftrightarrow \left\langle \underbrace{\sqrt{\beta} \mathbb{1}(u_l(x) = S)}_{\phi(x)}, \underbrace{\sqrt{\beta} \mathbb{1}(u_l(z) = S)}_{\phi(z)} \right\rangle_{l,S}$$

$$\begin{aligned} c) \quad k(x, x') &= \sum_{l=1}^2 \beta_l \sum_{S \in A^l} \mathbb{1}(u_{l,2}(x) = S) \mathbb{1}(u_{l,2}(x') = S) \\ &= \left\langle \mathbb{1}(u_{l,2}(x) = S) \right\rangle_{l,S} \left(\mathbb{1}(u_{l,2}(x') = S) \right)_{l,S} \end{aligned}$$

3 Fisher Kernel

$$\begin{aligned} a) \quad k(x, x') &= G_x^T E_z [G_z G_z^T]^{-1} G_{x'} \\ &= (\Sigma^{-1} (x - \mu))^T E_z [(\Sigma^{-1} (x - z)) (\Sigma^{-1} (x' - z))^T] \\ &\quad \Sigma^{-1} (x' - \mu) \\ &= (x - \mu)^T \Sigma^{-1} (\underbrace{E_z [\Sigma^{-1} (x - z) (\Sigma^{-1} (x' - z))^T]}_{\mathbb{1}}) \Sigma^{-1} (x' - \mu) \\ &= (x - \mu)^T \Sigma^{-1} (\underbrace{(\Sigma^{-1})^{-1} \Sigma^{-1}}_{= \mathbb{1}}) (x' - \mu) \\ &= (x - \mu)^T \Sigma^{-1} (x' - \mu) \end{aligned}$$

$$\begin{aligned} b) \quad k(x, x') &= (x - \mu)^T \Sigma^{-1} (x' - \mu) \\ &= (x - \mu)^T L L^T (x' - \mu) \\ &= \underbrace{(L^T (x - \mu))^T}_{\phi(x)} \underbrace{(L^T (x' - \mu))}_{\phi(x')} = \langle \phi(x), \phi(x') \rangle \end{aligned}$$