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

$$\underline{\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(x_i, x_j)} \geq 0 \quad \forall c.\ c^T K c \geqslant 0$$

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

$$\sum_i\sum_j c_i c_j \sum_{t=-\infty}^{\infty} \left([x_i * x_j]_t\right)^2 = \sum_i\sum_j c_i c_j \sum_t [x_i * x_j]_t [x_j * x_i]_t$$

$$= \sum_i\sum_j c_i c_j \sum_t \sum_\tau x_i(\tau)\, x_j(t-\tau) \cdot \sum_{\tau'} x_j(\tau') \cdot x_i(t-\tau')$$

define $s = t - \tau - \tau' \iff t = s + \tau + \tau'$

$$= \sum_i\sum_j c_i c_j \sum_s \sum_\tau \sum_{\tau'} x_i(\tau)\, x_j(s+\tau')\, x_j(\tau')\, x_i(s+\tau)$$

$[x_i * x_i]_s$ "autocorrelation"

$$= \sum_i\sum_j c_i c_j \sum_s \Big[\sum_\tau x_i(\tau)\, x_i(s+\tau)\Big]\Big[\sum_{\tau'} x_j(\tau')\, x_j(s+\tau')\Big]$$

$$= \sum_s \Big[\sum_i c_i \sum_\tau x_i(\tau)\, x_i(s+\tau)\Big]\Big[\sum_j c_j \sum_{\tau'} x_j(\tau')\, x_j(s+\tau')\Big] = \sum_s \Big[\sum_i c_i \sum_\tau x_i(\tau)\, x_i(s+\tau)\Big]^2 \geqslant 0$$

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

$$k(x, x') = \sum_s \underbrace{\sum_\tau x(\tau)\, x(s+\tau)}_{} \cdot \underbrace{\sum_{\tau'} x'(\tau')\, x'(s+\tau')}_{}$$

$$= \sum_s \quad [x \star x]_s \qquad [x' \star x']_s$$

$$= \langle \underbrace{[x \star x]}_{\phi(x)}, [x' \star x'] \rangle$$

Exercise 2: Weighted Degree Kernels (20 P)

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

m-gram at location l

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

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.

$u_{43}(x)$

x	AAA$\boxed{\text{CAA}}$ATAAGTAACTAATCTTTTAGGAAGAACGTTTCAACCATTTTGAG
#1-mers	.|.|.|||.|..||.|.|..|||.||...|....|...|||......|..
#2-mers	.....||.....|.......||..|..............||..........
#3-mers	.....|...............|...|...............|..........
x'	TACCTAATTATGAAATTAAATTTCAGTGTGCTGATGGAAACGGAGAAGTC

(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 $c_1, \ldots, c_N$.

$$\sum_i\sum_j c_i c_j \sum_m \beta_m \sum_l I(u_{lm}(x_i) = u_{lm}(x_j))$$

$$= \sum_i\sum_j c_i c_j \sum_m \beta_m \sum_l \sum_{s \in \mathcal{A}^m} I(u_{lm}(x_i) = s) \cdot I(u_{lm}(x_j) = s)$$

$$= \sum_m \beta_m \Big[\sum_i c_i \sum_l \sum_s I(u_{lm}(x_i) = s)\Big]^2 \geqslant 0$$

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

$$k(x, x') = \sum_l \sum_{s \in \mathcal{A}} \beta \cdot I(u_l(x) = s)\, I(u_l(x') = s)$$

$$= \sum_l \sum_{s \in \mathcal{A}} \sqrt{\beta}\, I(u_l(x) = s) \cdot \sqrt{\beta}\, I(u_l(x') = s)$$

$$= \langle \underbrace{\left(\sqrt{\beta}\, I(u_l(x) = s)\right)_{ls}}_{\phi(x) \in \mathbb{R}^{L \cdot 4}}, \left(\sqrt{\beta}\, I(u_l(x') = s)\right)_{ls} \rangle$$

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

$$k(x, x') = \sum_l \sum_{s \in \mathcal{A}^2} \beta\, I(u_{l2}(x) = s)\, I(u_{l2}(x') = s)$$

$$= \langle \underbrace{\left(\sqrt{\beta}\, I(u_{l2}(x) = s)\right)_{ls}}_{\phi(x) \in \mathbb{R}^{(L-1) \cdot 16}}, \left(\sqrt{\beta}\, I(u_{l2}(x') = s)\right)_{ls} \rangle$$

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:

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

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:

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

$$-\log p_\theta(x) = \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) + \text{cst.}$$

$$G_x = \frac{\partial}{\partial \mu} -\log p_\theta(x) = \Sigma^{-1}(x-\mu)$$

$$k(x, x') = G_x^T \left[\mathbb{E}_{z \sim p_\theta(x)}[G_z G_z^T]\right]^{-1} G_{x'}$$

$$= (x-\mu)^T \Sigma^{-1} \cancel{\Big[\Sigma^{-1} \underbrace{\mathbb{E}_z[(z-\mu)(z-\mu)^T]}_{\cancel{\Sigma}} \cancel{\Sigma^{-1}}\Big]^{-1} \cancel{\Sigma^{-1}}} (x'-\mu)$$

$$= (x-\mu)^T \Sigma^{-1}(x'-\mu)$$

"Mahalanobis distance"

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

Cholesky decomposition: $\Sigma^{-1} = LL^T$

$$(x-\mu)^T L L^T (x'-\mu) = \langle \underbrace{L^T(x-\mu)}_{\phi(x)}, L^T(x'-\mu) \rangle$$

Eigendecomposition: $\Sigma^{-1} = V\Lambda V^T = V\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}V^T$