ML2 Sheet 4

(2) (a) $$k(x,y) = \sum_{m=1}^{M} \beta_m \sum_{\ell=1}^{L+1-m} I\left(u_{\ell,m}(x) = u_{\ell,m}(z)\right)$$

$$= \sum_{m=1}^{M} \beta_m \sum_{\ell=1}^{L+1-m} \prod_{k=\ell}^{\ell+m-1} \underbrace{I\left(x(k) = z(k)\right)}_{=:\, K_k(x,z)}.$$

It suffices to show that K_k is a kernel. ~~since products~~ (positive semi-definite)

~~sums and~~

$$\sum_{i,j=1}^{N} c_i c_j K_k(x_i, x_j) = \sum_{\substack{i,j=1 \\ x_i = x_j}}^{N} c_i c_j$$

$$= \sum_{S \in \mathcal{A}} \sum_{i,j \in M_S} c_i c_j ,$$

setting $M_S := \{ i \in \{1, \dots, N\} \mid x_i(k) = S \}$, for $S \in \mathcal{A}$ and fixed $k \in \{1, \dots, L\}$. This can be seen as the product $c^T A c$ for the vector $c := (c_1, \dots, c_N)^T$ and an appropriate matrix $A \in \mathbb{R}^{N \times N}$. ~~Let $P \in \mathbb{R}^{N \times N}$ be a permutation matrix that "orders" the indices c in a way so that for $\tilde{c} := Pc$ it holds that $c_1, \dots, c_{n_1} \in$ $1, \dots, n_1 \in M_G$, $n_1+1, \dots, n_2 \in M_A$, $n_2+1, \dots, n_3 \in M_T$ and $n_3+1, \dots, N \in M_C$.~~ Assume for a moment that the indices $1, \dots, N$ are "ordered" in a way so that $M_G = \{1, \dots, n_1\}$, $M_A = \{n_1+1, \dots, n_2\}$, $M_T = \{n_2+1, \dots, n_3\}$ and $M_C = \{n_3+1, \dots, N\}$. (*)

The matrix A in this case must have the form

$$A = \begin{pmatrix} \begin{matrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{matrix} & & & O \\ & \begin{matrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{matrix} & & \\ & & \begin{matrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{matrix} & \\ O & & & \begin{matrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{matrix} \end{pmatrix}$$

where the size of the blocks containing ~~ones~~ ones corresponds to the size of the sets M_s, $s \in \mathcal{A}$, i.e., the number of indices i so that $x_i(k) = s$. The eigenvalues of a block diagonal matrix are the eigenvalues of the blocks. A matrix of the form

$\mathbb{1}_n := \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}$ has rank one because it is the outer product of the one-vector $\mathbb{1}_n := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$ with itself, hence at most one eigenvalue can be nonzero. This eigenvalue is $n \geq 0$ because

$\mathbb{1}_n \cdot \mathbb{1}_n = \mathbb{1}_n \underbrace{\mathbb{1}_n^T \mathbb{1}_n}_{=n} = n \cdot \mathbb{1}_n$. This shows that all eigenvalues of A are nonnegative, hence A is positive semi-definite.

In the general case we can "order" the indices ~~vector c by a multiplication with an appropriate permutation matrix $P \in \mathbb{R}^{N \times N}$~~ the indices with an appropriate permutation $\pi : \{1, .., N\} \to \{1, .., N\}$ so that (*) holds for $\{\pi(1), .., \pi(N)\}$. Let $P \in \mathbb{R}^{N \times N}$ be the corresponding permutation matrix, i.e., $Pc = \begin{pmatrix} c_{\pi(1)} \\ \vdots \\ c_{\pi(N)} \end{pmatrix}$. Using $P^T = P^{-1}$, we can write

$$c^T A c = (Pc)^T P A P^T (Pc)$$

and we can apply the argument from before to obtain that $\tilde{A} := PAP^T$ is p.s.d.. Since A and $\tilde{A}$ are similar by definition of $\tilde{A}$, they have the same eigenvalues and also A has to be s.p.d., showing that K_k is a p.s.d. kernel.

b) $\underline{M=1:}$

$$k(x,z) = \sum_{l=1}^{L} I(u_{l,1}(x) = u_{l,1}(z))$$

We can write

$$I(u_{l,1}(x) = u_{l,1}(z)) = \sum_{S\in A} x_{\{S\}}(x_l)\cdot x_{\{S\}}(z_l),$$

hence, a possible feature map is

$$\Phi: A^L \to \mathbb{R}^{4L}, \quad x \mapsto (x_{\{G\}}(x_1),\dots,x_{\{C\}}(x_1),\dots,x_{\{G\}}(x_L),\dots,x_{\{C\}}(x_L))^T$$
$$= (x_{\{S\}}(x_l))_{S\in A,\, l\in\{1,..,L\}}$$

(c) $\underline{M=2,\ \beta_1=0\ \beta_2=1:}$

$$k(x,z) = \sum_{l=1}^{L-1} I(u_{l,2}(x) = u_{l,2}(z)).$$

Writing

$$I(u_{l,2}(x) = u_{l,2}(z)) = \Big(\sum_{S\in A} x_{\{S\}}(x_l)\cdot x_{\{S\}}(z_l)\Big)\cdot\Big(\sum_{S\in A} x_{\{S\}}(x_{l+1})\cdot x_{\{S\}}(z_{l+1})\Big)$$
$$= \sum_{S,\hat{S}\in A} x_{\{S\}}(x_l)\, x_{\{\hat{S}\}}(x_{l+1})\cdot x_{\{S\}}(z_l)\, x_{\{\hat{S}\}}(z_{l+1}),$$

we see that a possible feature map is

$$\Phi: A^L \to \mathbb{R}^{16(L-1)}, \quad \cancel{x \mapsto (x_{\{S\}}}$$
$$x \mapsto (x_{\{G\}}(x_1)\, x_{\{G\}}(x_2),\ x_{\{G\}}(x_1)\cdot x_{\{A\}}(x_2),\dots,x_{\{C\}}(x_1)\cdot x_{\{C\}}(x_2),\dots$$
$$x_{\{G\}}(x_{L-1})\cdot x_{\{G\}}(x_L),\dots,x_{\{C\}}(x_{L-1})\cdot x_{\{C\}}(x_L))^T$$
$$= (x_{\{S\}}(x_l)\, x_{\{\hat{S}\}}(x_{l+1}))_{S,\hat{S}\in A,\, l\in\{1,..,L-1\}}$$

Notas

2) • $G_x = \frac{\partial}{\partial\mu} \log p_\mu(x)$

$$= \log p_\mu(x) \cdot \frac{\partial}{\partial\mu}\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

$$= \log p_\mu(x) \, \Sigma^{-1}(x-\mu)$$

3 (a) • $\frac{\partial}{\partial\mu} \log p_\mu(x) = \frac{\partial}{\partial\mu}\left(\log\left(\frac{1}{\sqrt{(2\pi)^d \det \Sigma}}\right) - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$

$$= \Sigma^{-1}(x-\mu) = G_x$$

• Let $z \sim p_\mu$. Note that $\Sigma^T = \Sigma$, hence $(\Sigma^{-1})^T = \Sigma^{-1}$

$$\mathbb{E}[G_z G_z^T] = \mathbb{E}\left[\Sigma^{-1}(z-\mu)(z^T-\mu^T)\Sigma^{-1}\right]$$

$$= \Sigma^{-1} \underbrace{\mathbb{E}\left[(z-\mu)(z^T-\mu^T)\right]}_{=\Sigma} \Sigma^{-1}$$

$$= \Sigma^{-1}$$

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

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

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

(b) Σ^{-1} is symmetric and positive definite, hence it possesses a ~~(unique)~~ Cholesky decomposition $\Sigma^{-1} = LL^T$. Set $\phi: \mathbb{R}^d \to \mathbb{R}^d$, $x \mapsto L^T(x-\mu)$, to get $\phi(x)^T \phi(x') = k(x,x')$.