Data Mining - Assignment 5 - Jennifer Probst - 16703423

Exercise 1

pol. kernel: $k(x,x') = (\langle x, x' \rangle + c)^p$

a) given: $x = (x_1, x_2)$, $x' = (x_1', x_2')$ with $p = 2$, $c = 1$

$$k(x, x') = (\langle (x_1, x_2), (x_1', x_2') \rangle + 1)^2 = (\langle (x_1, x_2), (x_1', x_2') \rangle^2 + 2\langle (x_1, x_2), (x_1', x_2') \rangle + 1$$

$$= x_1^2 x_1'^2 + x_2 x_2'^2 + 2x_1 x_1' x_2 x_2' + 2x_1 x_1' + 2x_2 x_2' + 1$$

$$= (x_1^2, x_2^2, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1) \cdot (x_1'^2, x_2'^2, \sqrt{2}x_1'x_2', \sqrt{2}x_1', \sqrt{2}x_2', 1)$$

$$= \phi(x) \cdot \phi(x')$$

→ $\phi(x)$ has dimensionality 6.

if we take $c = 0$ → $k(x, x') = (\langle x, x' \rangle)^p$

eg: $p = 2 \rightarrow (\langle x, x' \rangle)^p = x^2 + 2xx' + x'^2$

$p = 3 \rightarrow (\langle x, x' \rangle)^p = x^3 + 3x^2x' + 3xx'^2 + x'^3$

Pascal's triangle

	1	
	1 1	
$p=2$	1 2 1	dim = 3
$p=3$	1 3 3 1	dim = 4
$p=4$	1 4 6 4 1	dim = 5
	⋮	

→ we observe that the factors for the monomials when the polynomial kernel if factorised out, can be read of the pascal's triangle. For $p \geq 2$ we then find the dimension of the feature space to be the sum of the number of monomials / factors for monomials ⇒ dimensionality of ϕ scales with $\underline{p+1}$.

b) No, we don't need to represent the feature space explicitly, as we are only interested in the high-dim. relationships.

an example with small dimension (dim 3) and $p = 2$, $c = 1/2$

(1) $\boxed{k(a,b) = (\langle a, b \rangle + 1/2)^2} = (a, a^2, 1/2) \cdot (b, b^2, 1/2)$ (*)

→ to find the high-dim. relationship we don't need the explicit representation (*) as we can plug in values for a and b into (1) directly. Eg for $a = 4$ and $b = 1$ → $k(a,b) = (4 + 1/2)^2 = 20.25$

Exercise 3:

a) $k(x,x') = \underbrace{3\langle x, x'\rangle^4}_{a} + \underbrace{1}_{b} + \underbrace{x^Tx'}_{c} + \underbrace{\exp\left(-\frac{1}{2\sigma^2}\|x-x'\|^2\right)}_{d}$

a: $\langle x, x'\rangle$ is the linear kernel → $\langle x, x'\rangle^4$ product of kernels
→ $\lambda = 3$ → from (2b) → a is a kernel

b: constant all-ones kernel (slides):
$\phi(x) = 1 \rightarrow \langle \phi(x), \phi(x')\rangle = 1$

c: x^Tx' is also the linear kernel → $\langle x, x'\rangle = x^Tx'$

d: gaussian radial basis function kernel (slides)

→ also from lecture slides: sum of valid kernels → valid kernel.
a + b + c + d → valid kernel ☒

b) $k_{Gxy}(x, x') = \sum_{s\in S, s'\in S'} k_{base}(s, s')$ with S and S' being the sets of all k-mers of X and X' respectively starting with G.

we define: $k_{base}(s,s') := \sum_{j=1}^{3} k_{delta}(s_j, s_j')$

where s_{i1} stands for the first element of 3mer s_i etc.

c) we get $S_1 = \{(GPA), (GFA), (GPP), (GAD)\}$
$S_2 = \{(GDQ), (GPV), (GRT)\}$
$S_3 = \{(GFP)\}$

So: • $k_{Gxy}(X_1, X_2) = \sum_{s_1\in S_1, s_2\in S_2} k_{base}(s_1, s_2) = k_{base}(GPA, GDQ) + \ldots + k_{base}(GAD, GRT)$

$= \underbrace{1+2+1}_{k_{base}(GPA,\cdot)} + \underbrace{1+1+1}_{k_{base}(GFA,\cdot)} + \underbrace{1+2+1}_{k_{base}(GPP,\cdot)} + \underbrace{1+1+1}_{k_{base}(GAD,\cdot)} = 14$

• $k_{Gxy}(X_1, X_3) = \sum_{s_1\in S_1, s_3\in S_3} k_{base}(s_1, s_3) = 1 + 2 + 2 + 1 = 6$

d) sequences not a multiple of 3 should not be a problem as we scan all triplets and look independently if they start with a 'G'.
A problem can be that longer sequences have higher similarities as they contain more 3-mers starting with a 'G' on average.
So if you compare sequences with unequal length you will have to account for the length otherwise the comparison will not be meaningful.

Exercise 2:

a) $X \subset \mathbb{R}^d$, prove linear kernel is kernel: (*) change order of summation

→ show: $\forall n \in \mathbb{N}$ and $\forall \{x_1, .., x_n\} \in \mathbb{R}^d$

→ Gram mat K is pos definite: $\forall c_i \in \mathbb{R}: \sum_{i,j} c_i c_j K_{ij} \geq 0$

↳ linear kernel: $K_{ij} = \langle x_i, x_j \rangle = \sum_{a=1}^{d} x_{ia} x_{ja}$

$$\sum_{i,j} c_i c_j K_{ij} = \sum_{i,j} c_i c_j \langle x_i, x_j \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \sum_{a=1}^{d} x_{ia} x_{ja}$$

$$\stackrel{(*)}{=} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{d} c_i c_j x_{ia} x_{ja} \stackrel{(*)}{=} \sum_{a=1}^{d} \left(\sum_{i=1}^{n} c_i x_{ia}\right)\left(\sum_{j=1}^{n} c_j x_{ja}\right)$$

$$= \sum_{a=1}^{d} \left(\sum_{i=1}^{n} c_i x_{ia}\right)^2 \geq 0 \quad \text{as } x^2 \geq 0 \ \forall x \in \mathbb{R}$$

∎

rename var j to i

b) $X \subset \mathbb{R}^d$, prove that dot product in any feature space is a kernel.

→ show: $\forall n \in \mathbb{N}$ and $\forall \{x_1, .., x_n\} \in \mathbb{R}^d$

→ Gram mat K is pos definite: $\forall c_i \in \mathbb{R}: \sum_{i,j} c_i c_j K_{ij} \geq 0$

↳ $K_{ij} = k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$

$$\rightarrow \sum_{i,j} c_i c_j K_{ij} = \sum_{i,j} c_i c_j \langle \phi(x_i), \phi(x_j) \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \sum_{a=1}^{b} \phi(x_{ia}) \phi(x_{ja})$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{b} c_i c_j \phi(x_{ia}) \phi(x_{ja}) = \sum_{a=1}^{b} \left(\sum_{i=1}^{n} c_i (\phi(x_{ia}))\right)\left(\sum_{j=1}^{n} c_j (\phi(x_{ja}))\right)$$

$$= \sum_{a=1}^{b} \left(\sum_{i=1}^{n} c_i (\phi(x_i))\right)^2 \geq 0 \quad \text{as } x^2 \geq 0 \ \forall x \in \mathbb{R}$$

∎

with $b = \dim \phi(x_i)$

c) $\sum_{i,j} c_i c_j K_{ij} = \sum_{i,j} c_i c_j k_3(x_i, x_j)$

i)
$$= \sum_{i,j} x_i x_j (k_1(x_i, x_j) + k_2(x_i, x_j))$$

$c_i, c_j \geq 0$ and k_1, k_2 valid kernels

$$= \underbrace{\sum_{i,j} c_i c_j k_1(x_i, x_j)}_{\geq 0} + \underbrace{\sum_{i,j} c_i c_j k_2(x_i, x_j)}_{\geq 0} \geq 0$$

ii) $\sum_{i,j} c_i c_j k_4(x_i, x_j) = \sum_{i,j} c_i c_j \lambda k_1(x_i, x_j) = \lambda^{2n} \underbrace{\sum_{i,j} c_i c_j k_1(x_i, x_j)}_{\geq 0}$

≥ 0 ∎

with $\lambda \in \mathbb{R}^+$, $c_i, c_j \geq 0$ and k_3 valid kernel