Homework 5

Ana Daria Laslo

Exercise 1

Exercise 1a

Given $x = (x_1, x_2)$ and $x' = (x'_1, x'_2)$, we subsitute into the kernel function for $c = 1$ and $p = 2$ and we obtain:

$$k(x, x') = x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2x_1 x'_1 x_2 x'_2 + 2x_1 x'_1 + 2x_2 x'_2 + 1$$

Therefore, $\phi(x)$ in this feature space is:

$$\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1 x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1)$$

,while the dimensionality of this feature space is equal to 6.

For $c = 0$ and p, the dimensionality of the feature space scales to p. Considering x and x' that are in a two-dimensional space $(d = 2)$, the exact number can be computed by analysing how many $d - 1$ partitions can be placed among p items. That would be equal to:

$$\begin{pmatrix} d + p - 1 \\ p \end{pmatrix} = \begin{pmatrix} p + 1 \\ p \end{pmatrix} = \frac{(p+1)!}{p!(p+1-p)!} = p + 1$$

Exercise 1b

No, there is no need to explicitly represent the feature space for non-linear kernels when using an SVM classifier. By using the dual formulation of the SVM and implementing the kernel trick, the inner products are calculated in an inner-product space using the kernel function, while avoiding the mapping. Even if the dimensionality is very high this allows efficient computation as the data is accessed only in terms of the inner products.

Exercise 2

Exercise 2a

To prove that the linear kernel is a kernel we have to show that the Gram matrix is positive definite. Let K be the Gram matrix (n x n)with respect to $x_1, x_2, \ldots, x_n$:

$$K_{ab} = k(x_a, x_b)$$

The matrix K is positive semi-definite if for all $c_a \mathbb{R}$, K satisfies:

$$\sum_{a,b}^{n} c_a c_b K_{ab} \geq 0$$

For our kernel, the expression for the Gram matrix can be written as follows:

$$\sum_{a,b}^{n} c_a c_b k(x_a, x_b) =$$

$$= \sum_{a,b}^{n} c_a c_b \langle x_{ai}, x_{bi} \rangle =$$

$$= \sum_{a,b}^{n} \langle c_a x_a, c_b x_b \rangle =$$

$$= \sum_{a,b}^{n} \sum_{i=1}^{d} c_a x_{ai} c_b x_{bi}$$

Reodering the summations we obtain:

$$\sum_{i=1}^{d} \sum_{a=1}^{n} \sum_{b=1}^{n} c_a x_{ai} c_b x_{bi} =$$

$$= \sum_{i=1}^{d} \sum_{a=1}^{n} c_a x_{ai} \sum_{b=1}^{n} c_b x_{bi}$$

Considering a and b are dummy variables over the same interval, one of them can be dropped and just keep one of them, resulting in:

$$\sum_{i=1}^{d} (\sum_{a=1}^{n} c_a x_{ai})^2 \geq 0$$

Summing multiple positive terms leads to a positive results which proves that the Gram matrix for the linear kernel is positive semi-definite.

Exercise 2b

To prove the dot product is a kernel in any feature space we can apply the same approach as used in *Exercise 2a*, replacing x_a, x_b with $\phi(x_a), \phi(x_b)$ in the expression, obtaining:

$$\sum_{a,b}^{n} c_a c_b k(x_a, x_b) =$$

$$= \sum_{a,b}^{n} c_a c_b \langle \phi(x_{ai}), \phi(x_{bi}) \rangle =$$

$$= \sum_{i=1}^{d} \sum_{a=1}^{n} c_a \phi(x_{ai}) \sum_{b=1}^{n} c_b \phi(x_{bi})$$

Applying the same method, we arrive at:

$$\sum_{i=1}^{d} (\sum_{a=1}^{n} c_a \phi(x_{ai}))^2 \geq 0$$

Exercise 2c

Part 1:

$$k_3(x, x') = k_1(x, x') + k_2(x, x')$$

To prove that k_3 is a valid kernel function we have to prove that its Gram matrix is positive semi-definite.

$$\sum_{a,b}^{n} c_a c_b k_3(x_a, x_b) =$$

$$= \sum_{a,b}^{n} c_a c_b (k_1(x_a, x_b) + k_2(x_a, x_b)) =$$

$$= \sum_{a,b}^{n} c_a c_b k_1(x_a, x_b) + \sum_{a,b}^{n} c_a c_b k_2(x_a, x_b)$$

This expression is equal to the summation of the Gram matrix of the two kernels. Summing two positive semi-definite Gram matrices results in a positive semi-definite matrix. Therefore k_3 is a kernel.

Part 2:

$$k_4(x, x') = \lambda k_1(x, x')$$

To prove that k_3 is a valid kernel function we have to prove that its Gram matrix is positive semi-definite.

$$\sum_{a,b}^{n} c_a c_b k_4(x_a, x_b) =$$

$$= \sum_{a,b}^{n} c_a c_b \lambda k_1(x_a, x_b) =$$

$$= \lambda \sum_{a,b}^{n} c_a c_b k_1(x_a, x_b) =$$

As λ is a positive scalar, and the sum is the Gram matrix of k_1, the expression is a positive semi-definite Gram matrix. Therefore k_4 is a kernel.

Exercise 3

Exercise 3a

To show that the given kernel is a kernel, we make use of the closure properties of kernels and some proeminent, defined kernels (lecture slides):

- $\langle x, x' \rangle^4$ is a polynomial kernel where c=0 and p=4, when multiplied by the positive scalar 3, $3\langle x, x' \rangle^4$ is still a kernel (see exercise 2)
- 1 is a valid kernel (the all ones kernel)
- $x^T x'$ is a valid kernel (exercise 2, linear kernel)
- $exp(-\frac{1}{2\sigma^2}||x - x'||^2)$ is a valid kernel (Gaussian Radial Basis Function kernel)

According to the closure property of kernels that the sum of kernels is a kernel, we can observe that summing the above listed terms we obtain a valid kernel.

Exercise 3b

The $k_{base}(s, s')$ kernel can be defined as following:

$$k_{base}(s, s') = \begin{cases} 0 & \text{if } s_1 \neq G \text{ } or \text{ } s'_1 \neq G \\ \sum_{i=1}^{3} k(s_i, s'_i) & \text{otherwise} \end{cases}$$

where $k(s_i, s'_i)$ is the delta (Dirac) kernel defined as following:

$$k(s_i, s'_i) = \begin{cases} 1 & \text{if } s_i = s'_i \\ 0 & \text{otherwise} \end{cases}$$

Exercise 3c

Calculating $k_{GXY}(X_1, X_2)$

In X_1 we can identify 4 3-mers that have the intial amino acid equal to G, for all other 3-mers the k_{base} would have been 0. In X_2 we can identify 3 3-mers. For all possible pairs between the two we will compute the result of the k_{base} and then summ them:

- GPA and GDQ = 1
- GPA and GPV = 2
- GPA and GRT = 1
- GFA and GDQ = 1
- GFA and GPV = 1
- GFA and GRT = 1
- GPP and GDQ = 1
- GPP and GPV = 2
- GPP and GRT = 1
- GAD and GDQ = 1
- GAD and GPV = 1
- GAD and GRT = 1

Therefore the result is 14.

Calculating $k_{GXY}(X_1, X_3)$

We apply the same approach as above. In X_1 we can identify 4 3-mers that have the intial amino acid equal to G, for all other 3-mers the k_{base} would have been 0. In X_3 we can identify one 3-mer: GFP. For all possible pairs between the two we will compute the result of the k_{base} and then summ them:

- GPA and GFP = 1
- GFA and GFP = 2
- GPP and GFP = 2
- GAD and GFP = 1

Therefore the result is 6.

Exercise 3d