XCS229i Problem Set 3

Due Monday, 24 August 2020.

Guidelines

1. These questions require thought, but do not require long answers. Please be as concise as possible.
2. If you have a question about this homework, we encourage you to post your question on our Slack channel, at `http://xcs229i-scpd.slack.com/`
3. Familiarize yourself with the collaboration and honor code policy before starting work.
4. For the coding problems, you may not use any libraries except those defined in the provided started code. In particular, ML-specific libraries such as `scikit-learn` are not permitted.

Submission Instructions

Written Submission: All students must submit an electronic PDF version of the written questions. We highly recommend typesetting your solutions via LaTeX, though it is not required. If you choose to hand write your responses, please make sure they are well organized and legible when scanned. The source LaTeXfor all problem sets is available on GitHub.

Coding Submission: All students must also submit a zip file of their source code. Create a submission using the following bash command:

(There is no zip submission for this problem set.)

If you are **NOT** able to successfully zip your code using the following bash command or do **NOT** have the zip command line tool on your machine, please run the following python script to zip your code as an alternative:

(There is no zip submission for this problem set.)

You should make sure to (1) restrict yourself to only using libraries included in the starter code, and (2) make sure your code runs without errors. Your submission will be evaluated by the auto-grader using a private test set and will be used for verifying the outputs reported in the writeup.

Honor code: We strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down the solutions independently, and without referring to written notes from the joint session. In other words, each student must understand the solution well enough in order to reconstruct it by him/herself. In addition, each student should write on the problem set the set of people with whom s/he collaborated. Further, because we occasionally reuse problem set questions from previous years, we expect students not to copy, refer to, or look at the solutions in preparing their answers. It is an honor code violation to intentionally refer to a previous year's solutions.

1. **[40 points] Constructing kernels**

In class, we saw that by choosing a kernel $K(x, z) = \phi(x)^T \phi(z)$, we can implicitly map data to a high dimensional space, and have a learning algorithm (e.g SVM or logistic regression) work in that space. One way to generate kernels is to explicitly define the mapping ϕ to a higher dimensional space, and then work out the corresponding K.

However in this question we are interested in direct construction of kernels. I.e., suppose we have a function $K(x, z)$ that we think gives an appropriate similarity measure for our learning problem, and we are considering plugging K into the SVM as the kernel function. However for $K(x, z)$ to be a valid kernel, it must correspond to an inner product in some higher dimensional space resulting from some feature mapping ϕ. Mercer's theorem tells us that $K(x, z)$ is a (Mercer) kernel if and only if for any finite set $\{x^{(1)}, \dots, x^{(n)}\}$, the square matrix $K \in \mathbb{R}^{n\times n}$ whose entries are given by $K_{ij} = K(x^{(i)}, x^{(j)})$ is symmetric and positive semidefinite. You can find more details about Mercer's theorem in the notes, though the description above is sufficient for this problem.

Now here comes the question: Let K_1, K_2 be kernels over $\mathbb{R}^d \times \mathbb{R}^d$, let $a \in \mathbb{R}^+$ be a positive real number, let $f : \mathbb{R}^d \mapsto \mathbb{R}$ be a real-valued function, let $\phi : \mathbb{R}^d \to \mathbb{R}^p$ be a function mapping from $\mathbb{R}^d$ to $\mathbb{R}^p$, let K_3 be a kernel over $\mathbb{R}^p \times \mathbb{R}^p$, and let $p(x)$ a polynomial over x with *positive* coefficients.

For each of the functions K below, state whether it is necessarily a kernel. If you think it is, prove it; if you think it isn't, give a counter-example.

[**Hint:** For part (e), the answer is that K *is* indeed a kernel. You still have to prove it, though. (This one may be harder than the rest.) This result may also be useful for another part of the problem.]

(a) **[5 point(s) Written]** $K(x, z) = K_1(x, z) + K_2(x, z)$

Since K_1 and K_2 are valid kernels, we know that these are symmetric and Positive Semi-Definite (PSD). This can be written as:

$$z^T K_1 z \geq 0$$

and

$$z^T K_2 z \geq 0$$

If K has to be a valid kernel, then it has to be symmetric and $aKa^T \geq 0$ where a is an arbitrary vector.

First, we define the two kernels:

$$K_1(x, z) = \phi_1(x)^T \phi_1(z)$$
$$K_2(x, z) = \phi_2(x)^T \phi_2(z)$$

Therefore,

$$K(x, z) = \phi_1(x)^T \phi_1(z) + \phi_2(x)^T \phi_2(z)$$

so,

$$\begin{aligned}
a^T K a &= \sum_i \sum_j a_i K a_j \\
&= \sum_i \sum_j a_i (\phi_1(x^{(i)})^T \phi_1(x^{(j)}) + \phi_2(x^{(i)})^T \phi_2(x^{(j)})) a_j \\
&= \sum_i \sum_j a_i \phi_1(x^{(i)})^T \phi_1(x^{(j)}) a_j + \sum_i \sum_j a_i \phi_2(x^{(i)})^T \phi_2(x^{(j)}) a_j \\
&= z^T K_1 z + z^T K_2 z \geq 0
\end{aligned}$$

The sum of two symmetric kernels is symmetric as well. So K is a **valid** kernel.

(b) **[5 point(s) Written]** $K(x, z) = K_1(x, z) - K_2(x, z)$

Here, K is not a valid kernel.
This is shown in the following counter-example. First, I define some valid kernels:

$$K_1(x,z) = x^T z + c$$
$$K_2(x,z) = x^T z + (c+1)$$

So

$$K(x,z) = -1$$

Hereby

$$-z^T z = -\sum_i \sum_j z_i z_j <= 0$$

So this is not PSD, so this is **not** a valid kernel

(c) **[5 point(s) Written]** $K(x,z) = aK_1(x,z)$
First, we define the kernels:

$$K_1(x,z) = \phi_1(x)^T \phi_1(z)$$

and hereby

$$K(x,z) = aK_1 = a\phi_1(x)^T \phi_1(z)$$

So we have to proof that

$$z^T K z = a z^T K_1 z \geq 0$$

which is true for every $a \geq 0$ due to properties of K_1. And we know that scaling a symmetric matrix is symmetric as well, since this is elementwise multiplication.
So K is a **valid** kernel.

(d) **[5 point(s) Written]** $K(x,z) = -aK_1(x,z)$
First, we define the kernels:

$$K_1(x,z) = \phi_1(x)^T \phi_1(z)$$

and hereby

$$K(x,z) = -aK_1 = a\phi_1(x)^T \phi_1(z)$$

So we have to proof that

$$z^T K z = -a z^T K_1 z \geq 0$$

which isnt true for all $a \geq 0$. So K is **not** a **valid** kernel.

Counter-Example:

Setting the kernel matrix to the identity matrix (Which is symmetric and PSD):

$$K_1(x,z) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

So the new kernel matrix, K, is

$$K(x,z) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

This matrix is not PSD, like:

$$\begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} < 0$$

So K is **not** a **valid** kernel.

(e) **[5 point(s) Written]** $K(x, z) = K_1(x, z)K_2(x, z)$
We see that

$$K_{ij} = K_1(x^{(i)}, z^{(j)})K_2(x^{(i)}, x^{(j)})$$

So we have (here just using a vector c)

$$c^T Kc = \sum_i \sum_j c_i K_{ij} c_j =$$
$$= \sum_i \sum_j c_i \phi_1(x^{(i)})\phi_1(x^{(j)})\phi_2(x^{(i)})\phi_2(x^{(j)})c_j$$

Now we split the sums

$$\sum_i c_i \phi_1(x^{(i)})\phi_2(x^{(i)}) \sum_j \phi_1(x^{(j)})\phi_2(x^{(j)})c_j$$
$$\left(\sum_i c_i \phi_1(x^{(i)})\phi_2(x^{(i)})\right)^2 \geq 0$$

So K is a **valid** kernel.

(f) **[5 point(s) Written]** $K(x, z) = f(x)f(z)$
We want to show that

$$a^T Ka = \sum_i \sum_j a_i K a_j$$
$$= \sum_i \sum_j a_i f(x^{(i)})f(z^{(j)})a_j$$

We see that this product over the sum can be written as a square (which for sure is non-negative)

$$a^T Ka = \left(\sum_i a_i f(x^{(i)})\right)^2 \geq 0$$

So K is a **valid** kernel.

(g) **[5 point(s) Written]** $K(x, z) = K_3(\phi(x), \phi(z))$

We see that K_3 is a kernel of a mapped input. So we define the Kernel's mapping ρ

$$K_3(x,z) = \rho(x)^T\rho(z)$$

So we have

$$K_3(\phi(x),\phi(z)) = \rho(\phi(x))^T\rho(\phi(z))$$

Hereby,

$$\begin{aligned} z^T K_3 z &= \sum_i \sum_j z_i \rho(\phi(x^{(i)}))^T \rho(\phi(x^{(j)})) z_j \\ &= \sum_i \sum_j z_i (\rho \circ \phi)(x^{(i)})^T (\rho \circ \phi)(x^{(j)}) z_j \end{aligned}$$

(h) **[5 point(s) Written]** $K(x,z) = p(K_1(x,z))$

Using p as a polynomial function (with positive coefficients), K is a linear combination of the Kernel K_1) of n degress:

$$\sum_{i=0}^{n} a_i (K_1)^i$$

Using (e) that the product of valid kernels (here K_1^n is the product of itself n times) also is a valid kernel, we see that the exponential terms in the polynomial are valid kernels.

And using (b) that valid kernels times a positive coefficient (p only has positive coefficients, a_i) also are valid kernels. So K is a sum of valid kernels.

Lastly, using (a) that the sum of 2 or more valid kernels is also a valid kernel, K is indeed a valid kernel.

So K is a **valid** kernel.