XCS229i Problem Set 3

Due Monday, 24 August 2020.

I formed study session with Ketki Ambekar.

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 $\{x^{(1)}, \ldots, x^{(n)}\}$, 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]** $K(x,z) = K_1(x,z) + K_2(x,z)$

Yes. K_1 and K_2 are valid symmetric and positive semidefinite kernels. As a result, the sum of K_1 and K_2 will yield a valid symmetric and positive semidefinite kernel.

$$
\begin{aligned}
&K(x,z)\\
&\quad = z^T K z\\
&\quad = z^T K_1 z + z^T K_2 z\\
&\quad \geq 0
\end{aligned}
$$

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

No. K_1 and K_2 are valid symmetric and positive semidefinite kernels. The resultant kernel K from the difference of K_1 and K_2 is not a valid kernel, because the difference may invalidate the positive and semidefinite requirement. Counter example:As seen below, $z^T K_2 z > z^T K_1 z$ will yield negative. For example, if $K_2 = 1.1K_1$, then $K(x,z) = z^T K_1 z - z^T K_2 z = -0.1 z^T K_1 z < 0$; hence not meeting the positive semidefinite requirement of a Mercer kernel.

$$
\begin{aligned}
&K(x,z)\\
&\quad = z^T K z\\
&\quad = z^T K_1 z - z^T K_2 z\\
&\quad \geq 0\\
z^T K_1 z &\geq z^T K_2 z
\end{aligned}
$$

(c) **[5 point(s) Written]** $K(x,z) = aK_1(x,z)$

Yes. a is a positive real number. K_1 is a valid symmetric and positive semidefinite kernel. Their product will yield a valid symmetric and positive semidefinite kernel.

$$
\begin{aligned}
&K(x,z)\\
&\quad = z^T K z\\
&\quad = a z^T K_1 z\\
&\quad \geq 0
\end{aligned}
$$

(d) **[5 point(s) Written]** $K(x,z) = -aK_1(x,z)$

No. a is a positive real number. K_1 is a valid symmetric and positive semidefinite kernel. Having a negative in front of a will yield negative; hence not meeting the positive semidefinite requirement of a Mercer kernel. Counter example: $a = 1$.

$$
\begin{aligned}
&K(x,z)\\
&\quad = z^T K z\\
&\quad = -a z^T K_1 z\\
&\quad < 0
\end{aligned}
$$

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

Yes, $K_1(x,z)K_2(x,z)$ is a valid symmetric and positive semidefinite kernel. The Gram matrix is given by $K = K_1 \circ K_2$, where $\circ$ is the Hadamard product.

$$
\begin{aligned}
K &= \sum_i \sum_j (\lambda_i u_i u_i^T) \circ (\mu_j v_j v_j^T)\\
&= \sum_i \sum_j \lambda_i \mu_j (u_i \circ v_j)(u_i \circ v_j)^T\\
&= \sum_k \gamma_k w_k w_k^T
\end{aligned}
$$

Thus,

$$
\begin{aligned}
K(x,z) &= z^T K z\\
&= \sum_k \gamma_k z^T w_k w_k^T z\\
&= \sum_k \gamma_k (w_k^T z)^2\\
&\geq 0
\end{aligned}
$$

(f) **[5 point(s) Written]** $K(x,z) = f(x)f(z)$

Yes, $f(x)f(z)$ is a valid symmetric and positive semidefinite kernel.

$$\begin{aligned} K(x,z) &= z^T K z \\ &= \sum_k z^T f(x^{(k)}) f(x^{(k)}) z \\ &= \sum_k (f(x^{(k)} z)^2 \\ &\geq 0 \end{aligned}$$

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

Yes, $K_3(\phi(x), \phi(z))$ is a valid symmetric and positive semidefinite kernel. The problem states that K_3 is a valid kernel. K will be a valid kernel as long as the mapping ϕ is in the same $\mathbb{R}$ of the kernel K_3.

$$\begin{aligned} K(x,z) &= z^T K z \\ &= \sum_k z^T K_3(\phi(x^{(k)}), \phi(x^{(k)})) z \\ &\geq 0 \end{aligned}$$

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

Yes, $p(K_1(x,z))$ is a valid symmetric and positive semidefinite kernel.$p(x)$ a polynomial over x with *positive* coefficients. From the previous solutions, we see that (a) sum of valid kernels yields a valid kernel, (c) multiplying a valid kernel with a positive coeficient yields a valid kernel, (e) multiplying valid kernels together yields a valid kernel. Thus, $K(x,z) = p(K_1(x,z))$ is a valid kernel.