

## Week 5 Lab: Concept Check Exercises

### Kernels

1. Fix  $n > 0$ . For  $x, y \in \{1, 2, \dots, n\}$  define  $k(x, y) = \min(x, y)$ . Give an explicit feature map  $\varphi : \{1, 2, \dots, n\}$  to  $\mathbb{R}^D$  (for some  $D$ ) such that  $k(x, y) = \varphi(x)^T \varphi(y)$ .

*Solution.* Define  $\varphi(x) = (\mathbf{1}(x \leq 1), \mathbf{1}(x \leq 2), \dots, \mathbf{1}(x \leq n))$ . Then  $\varphi(x)^T \varphi(y) = \min(x, y)$ .

2. Show that  $k(x, y) = (x^T y)^4$  is a positive semidefinite kernel on  $\mathbb{R}^d \times \mathbb{R}^d$ .

*Solution.*  $k_1(x, y) = x^T y$  is a psd kernel, since  $x^T y$  is an inner product on  $\mathbb{R}^d$ . Using the product rule for psd kernels, we see that

$$k(x, y) = k_1(x, y)k_1(x, y)k_1(x, y)k_1(x, y) = k_1(x, y)^4$$

is psd as well.

3. Let  $A \in \mathbb{R}^{d \times d}$  be a positive semidefinite matrix. Prove that  $k(x, y) = x^T A y$  is a positive semidefinite kernel.

*Solution.* Fix  $x_1, \dots, x_n \in \mathbb{R}^d$  and let  $X$  denote the matrix that has  $x_i^T$  as its  $i$ th row. Then note that  $(XAX^T)_{ij} = x_i^T A x_j = k(x_i, x_j)$ . Thus we are done if we can show  $XAX^T$  is positive semidefinite. But note that

$$\alpha^T XAX^T \alpha = (X^T \alpha)^T A (X^T \alpha) \geq 0,$$

since  $A$  is positive semidefinite.

4. Consider the objective function

$$J(w) = \|Xw - y\|_1 + \lambda \|w\|_2^2.$$

Assume we have a positive semidefinite kernel  $k$ .

- (a) What is the kernelized version of this objective?
- (b) Given a new test point  $x$ , find the predicted value.

*Solution.*

- (a)  $J(\alpha) = \|K\alpha - y\|_1 + \lambda \alpha^T K \alpha$ , where  $K_{ij} = k(x_i, x_j)$ . Here  $x_i^T$  is the  $i$ th row of  $X$ .
- (b)  $f_\alpha(x) = \sum_{i=1}^n \alpha_i k(x_i, x)$ .

5. Show that the standard 2-norm on  $\mathbb{R}^n$  satisfies the parallelogram law.

*Solution.*

$$\begin{aligned}\|x - y\|_2^2 + \|x + y\|_2^2 &= (\|x\|_2^2 - 2x^T y + \|y\|_2^2) + (\|x\|_2^2 + 2x^T y + \|y\|_2^2) \\ &= 2\|x\|_2^2 + 2\|y\|_2^2.\end{aligned}$$

6. Suppose you are given an training set of distinct points  $x_1, x_2, \dots, x_n \in \mathbb{R}^n$  and labels  $y_1, \dots, y_n \in \{-1, +1\}$ . Show that by properly selecting  $\sigma$  you can achieve perfect 0-1 loss on the training data using a linear decision function and the RBF kernel.

*Solution.* By selecting  $\sigma$  sufficiently small (say, much smaller than  $\min_{i \neq j} \|x_i - x_j\|_2$ ) we can use  $\alpha_i = y_i$  and get very pointy spikes at each data point. [Note: This is not possible if any repeated points have different labels, which is not unusual in real data.]

7. Suppose you are performing standard ridge regression, which you have kernelized using the RBF kernel. Prove that any decision function  $f_\alpha(x)$  learned on a training set must satisfy  $f_\alpha(x) \rightarrow 0$  as  $\|x\|_2 \rightarrow \infty$ .

*Solution.* Since  $f_\alpha(x) = \sum_{i=1}^n \alpha_i k(x_i, x)$  we have

$$\lim_{\|x\|_2 \rightarrow \infty} f_\alpha(x) = \lim_{\|x\|_2 \rightarrow \infty} \sum_{i=1}^n \alpha_i \exp\left(-\frac{\|x_i - x\|_2^2}{2\sigma^2}\right) = \sum_{i=1}^n \alpha_i \lim_{\|x\|_2 \rightarrow \infty} \exp\left(-\frac{\|x_i - x\|_2^2}{2\sigma^2}\right) = 0.$$

8. Consider the standard (unregularized) linear regression problem where we minimize  $L(w) = \|Xw - y\|_2^2$  for some  $X \in \mathbb{R}^{n \times m}$  and  $y \in \mathbb{R}^n$ . Assume  $m > n$ .

- Let  $w^*$  be one minimizer of the loss function  $L$  above. Give an infinite set of minimizers of the loss function.
- What property defines the minimizer given by the representer theorem (in terms of  $X$ )?

*Solution.*

- $\{w^* + v \mid v \in \text{null}(X)\}$ . Using the standard inner product on  $\mathbb{R}^n$ , we can also write  $\text{null}(X)$  as the set of all vectors orthogonal to the row space of  $X$ .
- $w^*$  lies in the row space of  $X$ .