CS 189 Introduction to Machine Learning
Summer 2018 DIS4

1 Kernels

For a function $k(\mathbf{x}_i, \mathbf{x}_j)$ to be a valid kernel, it suffices to show either of the following conditions is true:

1. k has an inner product representation: $\exists\ \Phi : \mathbb{R}^d \rightarrow \mathcal{H}$, where $\mathcal{H}$ is some (possibly infinite-dimensional) inner product space such that $\forall \mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d,\ k(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$.

2. For every sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in R^d$, the Gram matrix

$$K = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & k(\mathbf{x}_i, \mathbf{x}_j) & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & \cdots & k(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix}$$

is positive semidefinite. For the following parts you can use either condition (1) or (2) in your proofs.

(a) Show that the first condition implies the second one, i.e. if $\forall \mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d,\ k(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$ then the Gram matrix K is PSD.

(b) Given two valid kernels k_a and k_b, show that their linear combination

$$k(\mathbf{x}_i, \mathbf{x}_j) = \alpha k_a(\mathbf{x}_i, \mathbf{x}_j) + \beta k_b(\mathbf{x}_i, \mathbf{x}_j)$$

is a valid kernel, where $\alpha \geq 0$ and $\beta \geq 0$.

(c) Given a valid kernel k_a, show that

$$k(\mathbf{x}_i, \mathbf{x}_j) = f(\mathbf{x}_i) f(\mathbf{x}_j) k_a(\mathbf{x}_i, \mathbf{x}_j)$$

is a valid kernel.

(d) Given a positive semidefinite matrix A, show that $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top A \mathbf{x}_j$ is a valid kernel.

(e) Show why $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top(\mathrm{rev}(\mathbf{x}_j))$ (where $\mathrm{rev}(x)$ reverses the order of the components in x) is *not* a valid kernel.

(f) In the kernel ridge regression problem in homework 4, one could reach the conclusion that when there is no normalization factor ($\lambda = 0$), the solution of kernel ridge regression can be computed by:

$$\underset{\alpha}{argmin}\left[\frac{1}{2}\alpha^T\mathbf{K}\alpha\right]$$

where $\mathbf{K} = \mathbf{X}\mathbf{X}^\mathbf{T}$ and $\mathbf{X} \in \mathbb{R}^{n\times d}$ is the kernelized feature matrix. In this case, why $\mathbf{K}$ is a valid kernel important? Assume that $\mathbf{K}$ is computed by applying a kernel function k on every sample pair: $k(\mathbf{x}_i, \mathbf{x}_j)$.

2 Multivariate Gaussians: A review

(a) Consider a two dimensional random variable $Z \in \mathbb{R}^2$. In order for the random variable to be jointly Gaussian, a necessary and sufficient condition is that

- Z_1 and Z_2 are each marginally Gaussian, and
- $Z_1|Z_2 = z$ is Gaussian, and $Z_2|Z_1 = z$ is Gaussian.

A second characterization of a jointly Gaussian RV Z is that it can be written as $Z = AX$, where X is a collection of i.i.d. standard normal RVs and $A \in \mathbb{R}^{2\times 2}$ is a matrix.

Note that the probability density function of a Gaussian RV is:

$$f(z) = exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)/\sqrt{(2\pi)^k|\Sigma|}$$

.

Let X_1 and X_2 be i.i.d. standard normal RVs. Let U denote a random variable uniformly distributed on $\{-1, 1\}$, independent of everything else. Verify if the conditions of the first characterization hold for the following random variables, and calculate the covariance matrix Σ_Z.

- $Z_1 = X_1$ and $Z_2 = X_2$.
- $Z_1 = X_1$ and $Z_2 = X_1 + X_2$. (Use the second characterization to argue joint Gaussianity.)
- $Z_1 = X_1$ and $Z_2 = -X_1$.
- $Z_1 = X_1$ and $Z_2 = UX_1$.

(b) Use the above example to show that two Gaussian random variables can be uncorrelated, but not independent. On the other hand, show that two uncorrelated, jointly Gaussian RVs are independent.

(c) With the setup above, let $Z = VX$, where $V \in \mathbb{R}^{2\times 2}$, and $Z, X \in \mathbb{R}^2$. What is the covariance matrix Σ_Z? Is this also true for a RV other than Gaussian?

(d) Use the above setup to show that $X_1 + X_2$ and $X_1 - X_2$ are independent. Give another example pair of linear combinations that are independent.

(e) Given a jointly Gaussian RV $Z \in \mathbb{R}^2$ with covariance matrix $\Sigma_Z = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}$, how would you derive the distribution of $Z_1 | Z_2 = z$?

Hint: The following identity may be useful

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -\frac{b}{c} & 1 \end{bmatrix} \begin{bmatrix} \left(a - \frac{b^2}{c}\right)^{-1} & 0 \\ 0 & \frac{1}{c} \end{bmatrix} \begin{bmatrix} 1 & -\frac{b}{c} \\ 0 & 1 \end{bmatrix}.$$

