EECS545 Lecture 8 Quiz Solutions

1. Which of the following are true statements about kernels? **Choose all options that apply**:

   (a) A machine learning algorithm can be kernelized if it does not need explicit access to the feature vectors and instead only requires access to inner products of the feature vectors.

   (b) A Gram/kernel matrix must be positive semidefinite.

   (c) The product of two kernel functions is still a kernel function.

   (d) The sum of two kernel functions is always a kernel function.

   Solution: (a), (b), (c), (d). All of them are true.

2. What is the purpose of the kernel trick?

   (a) To transform the problem from regression to classification.

   (b) To transform the data into a richer feature space without explicitly computing the feature vector.

   (c) To transform the problem from supervised to unsupervised learning.

   (d) To transform a linear regression model to SVM.

   Solution: (b) Please check the slide 10 of Lecture 8 for an example.

3. (True/False) For any two documents $\mathbf{x}$ and $\mathbf{z}$, define k($\mathbf{x}$, $\mathbf{z}$) to equal the number of unique words that occur in both $\mathbf{x}$ and $\mathbf{z}$ (i.e., the size of the intersection of the sets of words in the two documents $\mathbf{x}$ and $\mathbf{z}$). This function cannot be considered as a kernel.

   Solution: False. We can make k as kernel by setting $\phi(\mathbf{x})$ as a binary vector whose i-th entry is 1 when the document x contains the i-th word and 0 if it doesn't.

4. (True/False) $k(\mathbf{x}, \mathbf{y}) = k_1(\mathbf{x}, \mathbf{y}) - k_2(\mathbf{x}, \mathbf{y})$ is a kernel if k_1 and k_2 are valid kernels, and those kernels are defined as $k_1(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \diamondsuit \mathbf{y} + 1)^2$ and $k_2(\mathbf{x}, \mathbf{y}) = \mathbf{x} \diamondsuit \mathbf{y}$ (Please assume $\diamondsuit$ is a predefined operation that can lead k_1 and k_2 valid kernel. For example, $\diamondsuit(\mathbf{x}, \mathbf{y}) = 2\mathbf{x} + \frac{\mathbf{y}}{2}$ could work as a candidate).

   Solution: True. $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \diamondsuit \mathbf{y})^2 + \mathbf{x} \diamondsuit \mathbf{y} + 1$ with each term in the sum being a kernel.

5. (True/False) A Gaussian kernel will always have better classification performance at a testing time compared to a linear kernel.

Solution: False. Bad hyperparameter choice or small (finite) training data size setting could lead the kernel to overfit.