Homework 4: kernels

CSCI145/MATH166, Mike Izbicki

DUE:

Name:

If you complete this assignment in LaTeX, then you will receive +2 pts extra credit.

Problem 1. Prove or provide a counter example for the following claims:

1. (5 pts) There exists a matrix that has all positive eigenvalues but has a negative element.

2. (5 pts) Every matrix with positive entries has positive eigenvalues.

Problem 2. Prove or disprove each claim:

1. (5pts) If k is a kernel function, then ck is also a kernel function for all $c \in \mathbb{R}$.

2. (5pts) If k_1, k_2 are kernel functions, then $k_1 + k_2$ is also a kernel function.

3. (5pts) If k_1, k_2 are kernel functions, then $k_1 \cdot k_2$ is also a kernel function.

Problem 3. (20 pts) Recall that a function $k : \mathcal{X}^2 \to \mathbb{R}$ is a *kernel function* if there exists a function $\phi : \mathcal{X} \to \mathbb{R}^d$ such that $k(\mathbf{x}_1, \mathbf{x}_2) = \phi(\mathbf{x}_1)^T \phi(\mathbf{x}_2)$. Prove that if $\mathcal{X} = \mathbb{R}^c$, then k being a kernel function is equivalent to k being positive semi-definite (i.e. $k(\mathbf{x}, \mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{X}$). Don't forget to show that the implication goes both ways.

HINT: We basically proved this in class when we showed that every Gram matrix is positive semi-definite, and every positive semi-definite matrix is a Gram matrix. The proof for this problem follows the exact same structure.

Problem 4. Let $\mathcal{H}_k$ be the set of kernelized linear classifiers with kernel k. That is,

$$\mathcal{H}_k = \left\{ \mathbf{x} \mapsto \text{sign}\left(\sum_{i=1}^{n} \alpha_i k(\mathbf{x}, \mathbf{x}_i) \right) : \alpha \in \mathbb{R}^n \right\} \tag{1}$$

Let k_d denote the polynomial kernel of dimension d. That is,

$$k_d(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1{}^T \mathbf{x}_2 + c)^d. \tag{2}$$

1. (15 pts) Explain why the approximation error of $\mathcal{H}_{k_i}$ is less then or equal to the approximation error of $\mathcal{H}_{k_j}$ for all $i \leq j$.

2. (10 pts) How does increasing the dimension of the polynomial kernel affect the Bayes, estimation, approximation, and generalization errors?

Problem 5. You are given a two class classification problem. All data points $\mathbf{x}_i$ with class label $y_i = 1$ satisfy $\|\mathbf{x}_i\|_1 = 1$, and all data points $\mathbf{x}_i$ with class label $y_i = 0$ satisfy $\|\mathbf{x}_i\|_1 = 1 - \epsilon$ for some small value of $\epsilon > 0$.

1. (5pts) Calculate the Bayes error for this problem.

2. (5pts) Explain why kernelized logistic regression with the polynomial kernel of degree 2 has non-zero approximation error. (A picture explanation is acceptable, you do not need a formal proof.)

3. (10pts) Design a finite dimensional kernel so that a kernelized logistic regression model will have zero approximation error.

4. (10pts) The Gaussian kernel also has zero approximation error on this problem. Which kernel would be better suited to solving this problem, and why?