DSC 140A - Homework 06

Due: Wednesday, May 15

Instructions: Write your solutions to the following problems either by typing them or handwriting them on another piece of paper. Show your work or provide justification unless otherwise noted. If you write code to solve a problem, include the code by copy/pasting or as a screenshot. You may use `numpy`, `matplotlib` (or another plotting library), and any standard library module, but no other third-party libraries unless specified. Submit homeworks via Gradescope by 11:59 PM.

Problem 1.

In this problem, you will demonstrate that kernel ridge regression is equivalent to ridge regression performed in feature space by showing that they make the same predictions on a concrete data set.

We'll use the toy data set:

i	$\vec{x}^{(i)}$	y_i
1	$(0, 2)^T$	1
2	$(1, 0)^T$	1
3	$(0, -2)^T$	1
4	$(-1, 0)^T$	1
5	$(0, 0)^T$	-1

We will define

$$\vec{\phi}(\vec{x}) = (1, x_1^2, x_2^2, \sqrt{2}\, x_1, \sqrt{2}\, x_2, \sqrt{2}\, x_1 x_2)^T.$$

It turns out that $\kappa(\vec{x}, \vec{x}') = (1 + \vec{x} \cdot \vec{x}')^2$ is a kernel for $\vec{\phi}$.

You can write code to perform any and all calculations in this problem. If you do, please show your code.

a) Learn a prediction rule $H_1(\vec{x})$ by performing ridge regression in the 6-dimensional feature space and report the optimal parameter vector $\vec{w}$. Use a regularization parameter of $\lambda = 2$.

b) Given a new point, $\vec{x} = (1, 1)^T$, what is the prediction made by your ridge regressor? That is, what is $H_1(\vec{x})$? Show your calculations / code.

c) Compute the kernel matrix, K, for this data.

d) Learn a prediction function $H_2(\vec{x})$ by solving the kernel ridge regression dual problem; that is, by finding the optimal vector $\vec{\alpha}$. Recall that there is an exact solution: $\vec{\alpha} = (K + n\lambda I)^{-1}\vec{y}$. Report the $\vec{\alpha}$ that you find.

Note: the lecture slides had originally stated the solution without the n in $n\lambda I$. This was a typo and has been fixed.

e) Let $\vec{x} = (1, 1)^T$, as before. What does your kernel ridge regressor predict for this point? That is, what is $H_2(\vec{x})$?

Hint: $H_2(\vec{x})$ should be the same as $H_1(\vec{x})$.

Problem 2.

Let X be a continuous random variable, and let Y be a binary random variable. Suppose the class conditional densities are know to be:

$$p_X(x \mid Y = 0) = \begin{cases} 1/5, & \text{if } 0 \le x \le 2 \\ 1/3, & \text{if } 2 < x \le 3 \\ 1/15, & \text{if } 3 < x \le 7 \end{cases}$$

$$p_X(x \mid Y = 1) = \begin{cases} 1/6, & \text{if } 0 \le x \le 1 \\ 1/8, & \text{if } 1 < x \le 5 \\ 1/6, & \text{if } 5 < x \le 7 \end{cases}$$

Suppose also that $\mathbb{P}(Y = 0) = 0.4$ and $\mathbb{P}(Y = 1) = 0.6$.

For what values of $x \in [0, 7]$ will the Bayes classifier predict $y = 1$?

Problem 3.

Let X be a continuous random variable, and let Y be a random class label (1 or 0). Recall that the Gaussian probability density function (pdf) is given by:

$$\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Suppose $p_X(x \mid Y = 1)$ is a Gaussian pdf with $\mu = 2$ and $\sigma = 3$ and that $p_X(x \mid Y = 0)$ is also Gaussian with $\mu = 5$ and $\sigma = 3$. Suppose, too, that $\mathbb{P}(Y = 1) = \mathbb{P}(Y = 0) = \frac{1}{2}$.

Recall that the Bayes error is the probability that the Bayes classifier makes an incorrect prediction. What is the Bayes error for this distribution? Show your work.

Hint: you'll want some way to compute the area under a Gaussian. You can use the tables that appear in the back of your statistics book, or you can use something like `scipy.stats.norm.cdf`. We'll let you read the documentation to see how to use it, but it may be helpful to remember that if F is the cumulative density function for a distribution with density f, then $\int_a^b f(x)\,dx = F(b) - F(a)$.