DSC 140A - Homework 05

Due: Wednesday, February 14

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer. Homeworks are due via Gradescope at 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

$$\vec{\phi}(\vec{x}) = (1, x_1^2, x_2^2, \sqrt{2}\,x_1, \sqrt{2}\,x_2, \sqrt{2}\,x_1x_2)^T.$$

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 $\lambda = 2$.

Solution: All the code for the solution is at the bottom.

The optimal parameter vector $\vec{w}$ is given by

$$\vec{w} = \begin{pmatrix} 0.087 \\ 0.152 \\ 0.174 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

rounded to 3 decimal places.

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.

Solution: The code for the solution is at the bottom.

The prediction made by the ridge regressor is $H_1(\vec{x}) = 0.413$. Rounded to 3 decimal places.

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

Solution: The code for the solution is at the bottom.

The kernel matrix K is given by

$$K = \begin{pmatrix} 25 & 1 & 9 & 1 & 1 \\ 1 & 4 & 1 & 0 & 1 \\ 9 & 1 & 25 & 1 & 1 \\ 1 & 0 & 1 & 4 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

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.

Solution: THe code for the solution is at the bottom.

The optimal vector $\vec{\alpha}$ is given by

$$\begin{pmatrix} 0.022 \\ 0.076 \\ 0.022 \\ 0.076 \\ -0.109 \end{pmatrix}$$

rounded to 3 decimal places.

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 $H_1(\vec{x})$.

Solution: The code for the solution is at the bottom.

The prediction made by the kernel ridge regressor is $H_2(\vec{x}) = 0.413$ rounded to 3 decimal places. This is the same as the prediction made by the ridge regressor $H_1(\vec{x})$.

Code

```
from sklearn.linear_model import Ridge
import numpy as np
import math


def phi(x1, x2):
    return np.array([1, x1**2, x2**2, math.sqrt(2)*x1, math.sqrt(2)*x2, math.sqrt(2)*x1*x2])


X = np.array([
    [0, 2],
    [1, 0],
```

```python
    [0, -2],
    [-1, 0],
    [0, 0]
])
y = [1, 1, 1, 1, -1]

phi_x = np.array([phi(x1, x2) for x1, x2 in X])


n, m = phi_x.shape
I = np.identity(m)
w = np.linalg.solve(phi_x.T @ phi_x + n*2*I, phi_x.T @ y)
# a
print(np.round(w, 3))

new_x = np.array([1, 1])
phi_new_x = phi(new_x[0], new_x[1])
# b
print(np.round(w @ phi_new_x, 3))

def kernel_func(x, x_prime):
    return (1 + np.dot(x, x_prime))**2


K = np.array([[kernel_func(x, x_prime) for x_prime in X] for x in X])
# c
print(K)

lambda_ = 2 * len(X)
alpha = np.linalg.solve(K + lambda_ * np.identity(len(X)), y)
# d
print(np.round(alpha, 3))

H2 = alpha @ np.array([kernel_func(new_x, x) for x in X])
# e
print(H2)
```