DSC 140A - Homework 05

Due: Wednesday, February 12

Instructions: Write your solutions to the following problems either by typing them or handwriting them on another piece of paper or on an iPad/tablet. Show your work or provide justification unless otherwise noted; submissions that don't show work might lose credit. If you write code to solve a problem, include the code by copy/pasting or as a screenshot. You may use numpy, pandas, 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.

A IAT_EX template is provided at http://dsc140a.com, next to where you found this homework. Using it is totally optional, but encouraged if you plan to go to grad school. See this video for a quick introduction to IAT_EX .

Problem 1.

In this problem, you will derive the solution to the ridge regression optimization problem.

Recall that the ridge regression regularized risk function is

$$\tilde{R}(\vec{w}) = \frac{1}{n} \sum_{i=1}^{n} (\vec{\phi}(\vec{x}^{(i)}) \cdot \vec{w} - y_i)^2 + \lambda \|\vec{w}\|^2.$$

Here, $\vec{\phi}(\vec{x})$ is a feature map. Also recall that this risk function can be equivalently written in matrix-vector form as

$$\tilde{R}(\vec{w}) = \frac{1}{n} \|\Phi\vec{w} - \vec{y}\|^2 + \lambda \|\vec{w}\|^2,$$

where Φ is the design matrix; its *i*th row is $\vec{\phi}(\vec{x}^{(i)})$.

- **a)** Show that $\tilde{R}(\vec{w}) = \frac{1}{n} \left(\vec{w}^T \Phi^T \Phi \vec{w} 2\vec{w}^T \Phi^T \vec{y} + \vec{y}^T \vec{y} \right) + \lambda \vec{w}^T \vec{w}.$
- b) So far this quarter, we have seen a few vector calculus identities. For example, we know that $\frac{d}{d\vec{w}}(\vec{w}^T\vec{w}) = 2\vec{w}.$

Using these identities, show that

$$\frac{d}{d\vec{w}}\tilde{R}(\vec{w}) = \frac{1}{n}\left(2\Phi^T\Phi\vec{w} - 2\Phi^T\vec{y}\right) + 2\lambda\vec{w}.$$

c) Show that the minimizer of $\tilde{R}(\vec{w})$ is $\vec{w}^* = (\Phi^T \Phi + n\lambda I)^{-1} \Phi^T \vec{y}$.

Problem 2.

The data set linked below contains data for performing non-linear regression. The first column is x (the independent variable), and the second column is y (the dependent variable).

https://f000.backblazeb2.com/file/jeldridge-data/010-nonlinear_regression/data.csv

Plotting the data shows that there is a non-linear relationship between x and y:



a) Fit a function of the form $H(\vec{x}) = w_1\phi_1(\vec{x}) + w_2\phi_2(\vec{x}) + \ldots + w_{50}\phi_{50}(\vec{x})$, where each $\phi_i(\vec{x})$ is a Gaussian basis function. Your 50 Gaussian basis functions should be equally-spaced, with the first at $\mu_1 = -10$ and the last at $\mu_{50} = 10$. The width of each Gaussian should be $\sigma = 1$. You should not augment \vec{x} .

For your answer, report only w_1 (the first component of \vec{w}), and show your code.

- b) Plot the prediction function $H(\vec{x})$ that you trained in the previous part, on top of the data. Provide your plot and the code used to generate it.
- c) You should see that the prediction function $H(\vec{x})$ slighly overfits the data. Now perform ridge regression on the same data, using the same Gaussian basis functions. Choose the regularization parameter λ to reduce overfitting (you may do so by trial and error no need to perform cross-validation). For your answer, state λ and plot your new prediction function on top of the data. Also provide your code.