CSE 151A - Discussion 04

Problem 1.

Suppose you have a set of labeled data points and have solved for the linear prediction rule $H(\vec{x}) = w_0 + w_1 x_1 + w_2 x_2$ by minimizing the MSE.

a) Describe at a high level what it means when $w_0 = 0$.

b) You now realize that you accidentally skipped a data point during your calculation; is it 100% necessary for you to recompute $\vec{w}$? If not, explain why.

Solution:
a) w_0 is the bias term, and represents the intercept/output when all values in $\vec{x}$ are 0. When $w_0 = 0$, it means that the prediction rule passes through the origin.

b) If the new data point lies exactly on the current prediction rule, then there is no need to recompute $\vec{w}$. The error for this new data point is 0 because its label exactly matches its predicted output. In any other scenario, we would need to recompute $\vec{w}$.

Problem 2.

Here is a data set of sales figures from different stores (this is a small sample taken directly from the data used in class, so the final answer will be slightly different).

store #	sales	sq. ft	# competitors
1	398	4.3	4
2	347	3.6	6
3	161	2.6	13
4	464	4.7	3

a) Let's try to predict net sales from two variables: the square footage of the store and the number of competing stores in the area. Our model will be:

$$\text{net sales} \approx w_0 + w_1 \times \text{sq. ft} + w_2 \times \#\text{ competitors}$$

Solve for w_0, w_1, and w_2 and rewrite the above equation using these values.

HINTS:
You will need the design matrix, X, and the vector $\vec{y}$.
Then, you can solve $\vec{w} = (X^T X)^{-1} X^T \vec{y}$.

b) Suppose your friend wants to open a store, but can only afford to do so if they can make more than \$300 in sales. If there are 5 competitors in the area, how large (in sq. ft) must the store be in order for your friend to make ends meet?

Solution:

a) $\vec{w} = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 249.7 \\ 54.8 \\ -17.8 \end{pmatrix}$

$\boxed{\text{net sales} \approx 249.7 + 54.8 \times \text{sq. ft} - 17.8 \times \text{\# competitors}}$

The design matrix X and the observation vector $\vec{y}$ can be computed directly by taking the values from the given data table. From there, X^T should be easy to determine as well.

$$X = \begin{pmatrix} Aug(\vec{x}^{(1)}) \\ Aug(\vec{x}^{(2)}) \\ Aug(\vec{x}^{(3)}) \\ Aug(\vec{x}^{(4)}) \end{pmatrix} = \begin{pmatrix} 1 & 4.3 & 4 \\ 1 & 3.6 & 6 \\ 1 & 2.6 & 13 \\ 1 & 4.7 & 3 \end{pmatrix} \quad X^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 4.3 & 3.6 & 2.6 & 4.7 \\ 4 & 6 & 13 & 3 \end{pmatrix} \quad \vec{y} = \begin{pmatrix} 398 \\ 347 \\ 161 \\ 464 \end{pmatrix}$$

We will now use these matrices to solve for $\vec{w}$

$$(X^T X) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 4.3 & 3.6 & 2.6 & 4.7 \\ 4 & 6 & 13 & 3 \end{pmatrix} \begin{pmatrix} 1 & 4.3 & 4 \\ 1 & 3.6 & 6 \\ 1 & 2.6 & 13 \\ 1 & 4.7 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 15.2 & 26 \\ 15.2 & 60.3 & 86.7 \\ 26 & 86.7 & 230 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 186.17 & -36.40 & -7.33 \\ -36.40 & 7.15 & 1.42 \\ -7.33 & 1.42 & 0.30 \end{pmatrix}$$

$$\vec{w} = (X^T X)^{-1} X^T \vec{y} = \begin{pmatrix} 186.17 & -36.40 & -7.33 \\ -36.40 & 7.15 & 1.42 \\ -7.33 & 1.42 & 0.30 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 4.3 & 3.6 & 2.6 & 4.7 \\ 4 & 6 & 13 & 3 \end{pmatrix} \begin{pmatrix} 398 \\ 347 \\ 161 \\ 464 \end{pmatrix} = \begin{pmatrix} 249.7 \\ 54.8 \\ -17.8 \end{pmatrix}$$

b) $\boxed{\text{2.55 sq ft}}$

Recall our equation: net sales $\approx 249.7 + 54.8 \times$ sq. ft $- 17.8 \times$ # competitors.

We can turn this equation into an inequality, plug in values for net sales and # competitors, and then solve for sq. ft.

$300 \leq 249.7 + 54.8 \times \text{sq. ft} - 17.8(5)$

sq. ft $\geq \frac{50.3+17.8(5)}{54.8} = 2.55$

Problem 3.

Suppose we have 50 labeled data points and we wish to fit a linear regression model to them. We plan on running several experiments to decide on model complexity and determine a reasonable λ value for regularization. For each experiment, we will use Leave-One-Out Cross Validation to compute the overall error for that model.

It takes 5 seconds to compute the validation error for a single fold, and we want to evaluate 1-degree, 3-degree, 5-degree, 10-degree, and 20-degree polynomial models over 4 different values for λ. How many total seconds will it take for all of the experiments to finish?

Solution: $\boxed{\text{5000 seconds}}$

We are using Leave-One-Out Cross Validation, so we will have one fold per data point. Since it takes 5 seconds to compute the validation error for a single fold, this will result in 50*5=250 seconds to compute the overall validation error for each experiment.

We are optimizing our model over two parameters (polynomial degree and λ), so we will need to run a separate experiment for every combination of these values. There are 5 polynomial degree settings and

4 λ values, which yields 5*4=20 experiments.
With 20 experiments and 250 seconds needed per experiment, the total runtime is 20*250=5000 seconds (or 83.33 minutes).