2. (9 points) Consider the following data set with 4-dimensional data points (recall that each column represents one data point):

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad Y = \begin{bmatrix} 1.1 & 1.9 & 3.1 \end{bmatrix}$$

We perform ridge regression with a linear hypothesis class and no constant offset, i.e. $h(x^{(i)}; \Theta) = \theta^T x^{(i)}$.

(a) What is an optimal $\theta^*$ and its mean-squared error (MSE) for a minimizer of the ridge regression objective with $\lambda = 0$, on this data? (Note, $\theta^*$ may not be unique with $\lambda = 0$.)

**Solution:** There are many solutions because the number of features ($d = 4$) is larger than the number of data points ($n = 3$). The system of equations is:

$$\theta^T X = Y$$

One possible solution is:

$$\theta^* = [1, 0.1, -0.1, 0.1]^T$$

This can be seen by inspection, because $Y$ is essentially just the first feature in $X$ (the top row), with some corrections. The corrections are given by the one-hot encodings provided by the other features in $X$.

The MSE of this solution is zero.

(b) As $\lambda$ becomes very large, what will the MSE be of the $\theta^*$ that minimizes the ridge regression objective? It is OK to leave unsimplified, e.g. $5^2$.

**Solution:** As $\lambda$ becomes very large, $\hat{\theta}$ will become smaller and smaller. Eventually, $\hat{\theta} = 0$. This would lead to

$$\text{MSE} = \frac{1}{3}(1.1^2 + 1.9^2 + 3.1^2) = \frac{1}{3}(14.43) = 4.81.$$

(c) Each one of the following parameter vectors was obtained by minimizing the ridge regression objective with $\lambda = .01$, 1, and 100. Which was which? (We rounded to 3 decimals.)

$\theta = [0.789, 0.078, 0.081, 0.183]^T$

**Solution:** $\lambda = 1$

$\theta = [0.045, 0.004, 0.006, 0.010]^T$

**Solution:** $\lambda = 100$

$\theta = [0.945, 0.151, 0.010, 0.258]^T$

| |
|---|
| **Solution:** $\lambda = .01$ |