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

$\theta^* =$

MSE $=$

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

(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.)

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

$\lambda =$

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

$\lambda =$

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

$\lambda =$