Linear Regression *(find and work with a partner)*

In the case of *simple* linear regression, for each example we have a single feature x and associated response (output) variable y. Our model is:

$$h_{\boldsymbol{w}}(x) = w_0 + w_1 x$$

1. Name one goal of linear regression. Why would we want to fit a linear model?

2. In simple linear regression, how many features does each example have? How many parameters does our linear model have in this case?

   - # features (p) =

   - # model params =

3. What cost/loss function are we trying to minimize in simple linear regression?

4. *Small example.* Let $n = 2$ and $p = 1$, with the following data:

$$\boldsymbol{X} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

   Plot these two points – what should $\hat{w}_0$ and $\hat{w}_1$ be?

5. This week we derived the solution for simple linear regression:

$$\hat{w}_1 = \frac{\text{Cov}(\boldsymbol{x}, \boldsymbol{y})}{\text{Var}(\boldsymbol{x})} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \hat{w}_0 = \bar{y} - \hat{w}_1 \bar{x}$$

   where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. Use these equations to compute $\hat{w}_0$ and $\hat{w}_1$ and verify your answer to the previous question.

Linear Regression: Analytic Solution *(find and work with a partner)*

In the case of *multiple* linear regression, each example has p features. We still have an associated response (output) variable y for each example. We typically add a "fake 1" to the features of each example, so that $\vec{x} = [1\ x_1\ x_2 \cdots x_p]^T$. Our model is now the dot product between the weight vector $\vec{w}$ and the features:

$$h_{\vec{w}}(\vec{x}) = w_0 + w_1 x_1 + \cdots w_p x_p = \vec{w} \cdot \vec{x}$$

If we consider $\boldsymbol{X}$ to be the entire matrix of features (with dimensions $n \times (p+1)$) then we can reframe the linear regression problem in terms of vectors and matrices. In this case the analytic solution for the weight vector is:

$$\hat{w} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \vec{y}$$

1. Going back to our small example from the last page, we will use this matrix/vector form to verify our solution for $\vec{w}$ (even though $p = 1$ we can still use the multiple linear regression derivation). We again have the data: $(x_1, y_1) = (1, 0)$ and $(x_2, y_2) = (0, 1)$. First, rewrite $\boldsymbol{X}$ with the added column of 1's. Then write $\vec{y}$ as a vector.

2. Now use the matrix/vector analytic solution to verify the weight vector $\vec{w}$.