

Terminology recap

localhost:8888/nbclassic/notebooks/lec15-blank.ipynb#/slide-5-0

• Define the design matrix $X \in \mathbb{R}^{n \times 2}$, observation vector $\vec{y} \in \mathbb{R}^n$, and parameter vector $\vec{w} \in \mathbb{R}^2$ as follows:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$
vector of actual y values

$$\vec{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$\vec{y}_i = w_0 (i) + w_1 (x_i)$$





localhost:8888/nbclassic/notebooks/lec15-blank.ipynb#/slide-6-0





Rewriting mean squared error

• The mean squared error of the predictions in h is:

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - (w_0 + w_1 x_i))^2$$

$$\text{lquivden}^{\frac{1}{2}}$$

Equivalently, we have:

$$R_{\text{sq}}(\vec{w}) = \frac{1}{n} ||\vec{y} - \underbrace{X\vec{w}}_{\vec{h}}||^2$$



localhost:8888/nbclassic/notebooks/lec15-blank.ipynb#/slide-8-0



Minimizing mean squared error

- To minimize mean squared error, $R_{
 m sq}(ec{w})=rac{1}{n}\|ec{y}-Xec{w}\|^2$, we must choose the $ec{w}^*$ such that the error vector, $\vec{e} = \vec{y} - X\vec{w}^*$, is orthogonal to the columns of X.
- Equivalently, we have:

$$X^T \vec{e} = 0$$

What are the dimensions of $X^T \vec{e}$?

$$\frac{1}{2} \left(2 \times n \right) \times \left(n \times 1 \right)$$

$$= \left(2 \times 1 \right)$$

$$\begin{cases} e_1 \\ e_2 \\ = \begin{cases} y_1 - H(x_1) \\ y_2 - H(x_2) \\ \vdots \\ e_N \end{cases}$$



localhost:8888/nbclassic/notebooks/lec15-blank.ipynb#/slide-8-0

8

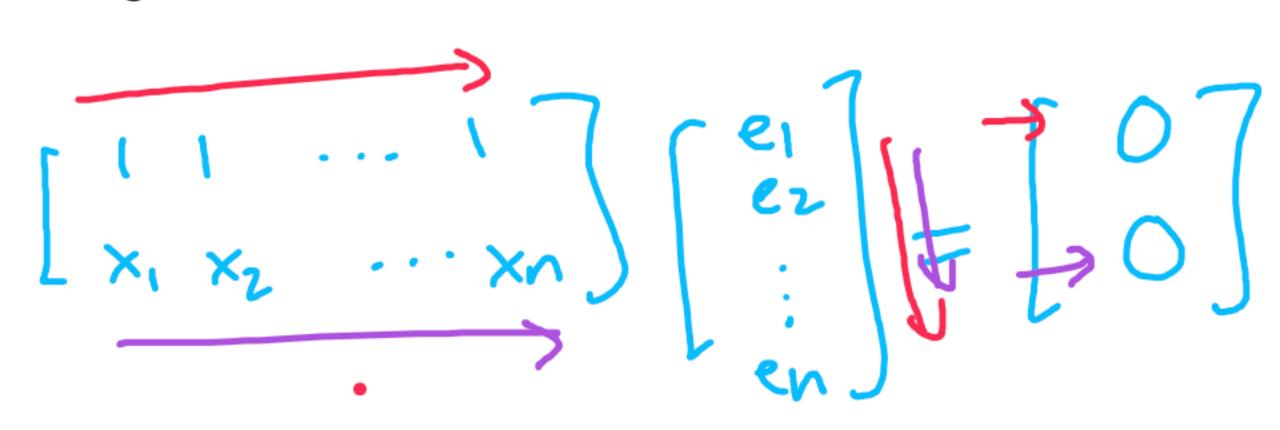
Minimizing mean squared error

- To minimize mean squared error, $R_{\rm sq}(\vec{w}) = \frac{1}{n} ||\vec{y} X\vec{w}||^2$, we must choose the \vec{w}^* such that the error vector, $\vec{e} = \vec{y} X\vec{w}^*$, is orthogonal to the columns of X.
- Equivalently, we have:

$$X^T \vec{e} = 0$$

What are the dimensions of $X^T \vec{e}$?

• Expanding, we have:

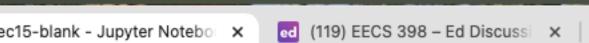


$$X^{T}(\vec{y} - X\vec{w}^{*}) = 0$$

$$X^{T}\vec{y} - X^{T}X\vec{w}^{*} = 0$$

$$X^{T}\vec{y} - X^{T}X\vec{w}^{*} = X^{T}\vec{y}$$





localhost:8888/nbclassic/notebooks/lec15-blank.ipynb#/slide-10-0



$$R_{\text{sq}}(w_0, w_1) = -\sum_{i=1}^{n} (y_i - (w_0 + w_1 x_i))^2$$

We found, using calculus, that:

$$\blacksquare w_0^* = \bar{y} - w_1^* \bar{x}$$



- Another way of finding optimal model parameters for simple linear regression is to find the \vec{w}^* that minimizes $R_{\rm sq}(\vec{w}) = \frac{1}{n} \|\vec{y} X\vec{w}\|^2$.
- The minimizer, if X^TX is invertible, is the vector $\vec{w}^* = (X^TX)^{-1}X^T\vec{y}$ If not, solve the **normal equations** from the previous slide.

Wed Mar 12 3:26 PM









The hypothesis vector

localhost:8888/nbclassic/notebooks/lec15-blank.ipynb#/slide-16-0

• When our hypothesis function is of the form:

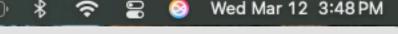
H(departure hour, day of month) = $w_0 + w_1 \cdot \text{departure hour} + w_2 \cdot \text{day of month}$

the hypothesis vector $\vec{h} \in \mathbb{R}^n$ can be written as:

$$\vec{h} = \begin{bmatrix} H(\text{departure hour}_1, \text{day}_1) \\ H(\text{departure hour}_2, \text{day}_2) \\ \dots \\ H(\text{departure hour}_n, \text{day}_n) \end{bmatrix} = \begin{bmatrix} 1 & \text{departure hour}_1 & \text{day}_1 \\ 1 & \text{departure hour}_2 & \text{day}_2 \\ \dots & \dots & \dots \\ 1 & \text{departure hour}_n & \text{day}_n \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$









The general solution

localhost:8888/nbclassic/notebooks/lec15-blank.ipynb#/slide-22-0

• Define the design matrix $X \in \mathbb{R}^{n \times (d+1)}$ and observation vector $\vec{y} \in \mathbb{R}^n$:

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(d)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(d)} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(d)} \end{bmatrix} = \begin{bmatrix} \operatorname{Aug}(\overrightarrow{x_1})^T \\ \operatorname{Aug}(\overrightarrow{x_2})^T \\ \vdots \\ \operatorname{Aug}(\overrightarrow{x_2})^T \end{bmatrix} \qquad \overrightarrow{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{\mathbf{N} \times \mathbf{N}}$$

