Lecture 12

## Multiple Linear Regression

DSC 40A, Fall 2025

Recap: Regression and linear algebra

#### Regression and linear algebra (Solution 1)

• 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:

$$X = egin{bmatrix} 1 & x_1 \ 1 & x_2 \ dots & dots \ 1 & x_n \end{bmatrix} \qquad ec{y} = egin{bmatrix} y_1 \ y_2 \ dots \ y_n \end{bmatrix} \qquad ec{w} = egin{bmatrix} w_0 \ w_1 \end{bmatrix}$$
 intercept

• How do we make the hypothesis vector,  $\vec{h}=X\vec{w}$ , as close to  $\vec{y}$  as possible? Use the parameter vector  $\vec{w}^*$ :

$$ec{w}^* = (X^TX)^{-1}X^Tec{y}$$

• Solution: We chose  $\vec{w}^*$  so that  $\vec{h}^* = X\vec{w}^*$  is the projection of  $\vec{y}$  onto the span of the columns of the design matrix, X and minimized the length of the projection error  $||\vec{e}|| = ||\vec{y} - X\vec{w}||$ .

#### Regression and linear algebra (Solution 2)

• 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:

$$egin{aligned} egin{aligned} egin{aligned\\ egin{aligned} egi$$

• How do we minimize the mean squared error  $R_{\rm sq}(\vec w)=\frac{1}{n}\|\vec y-X\vec w\|^2$  ? Using calculus the optimal paramter vector  $\vec w^*$  is:

$$ec{w}^* = (oldsymbol{X}^Toldsymbol{X})^{-1}oldsymbol{X}^Toldsymbol{ec{y}}$$

• Solution: we computed the gradient of  $R_{\rm sq}(\vec{w})$ , set it to zero and solved for  $\vec{w}$ .

## Multiple linear regression

	× <sup>(1)</sup>	x (2)	9	. d.
	departure_hour		minutes	commit
0	10.816667	15	68.0	
1	7.750000	16	94.0	
2	8.450000	22	63.0	
3	7.133333	23	100.0	
4	9.150000	30	69.0	

So far, we've fit **simple** linear regression models, which use only **one** feature ('departure\_hour') for making predictions.

#### The setup

• Suppose we have the following dataset.

ie following dataset.			comm	at the
	departure_hour	<b>7</b> – –	minutes	
row	×0	X X	<u></u>	
1	8.45	22	63.0	
2	8.90	28	89.0	
3	8.72	18	89.0	
_				

• We can represent each day with a **feature vector**,  $\vec{x}$ :

$$X_{1} = \begin{bmatrix} 7.45 \\ 22 \end{bmatrix}$$

$$Y_{1} = 6?$$

$$Y_{2} = 89$$

$$X_{2} = \begin{bmatrix} 8.30 \\ 22 \end{bmatrix}$$

#### Finding the optimal parameters

• To find the optimal parameter vector,  $\vec{w}^*$ , we can use the **design matrix**  $X \in \mathbb{R}^{n \times 3}$  and **observation vector**  $\vec{y} \in \mathbb{R}^n$ :

$$X = egin{bmatrix} 1 & \operatorname{departure\ hour}_1 & \operatorname{day}_1 \ 1 & \operatorname{departure\ hour}_2 & \operatorname{day}_2 \ \dots & \dots & \dots \ 1 & \operatorname{departure\ hour}_n & \operatorname{day}_n \end{bmatrix} \qquad ec{y} = egin{bmatrix} \operatorname{commute\ time}_1 \ \operatorname{commute\ time}_2 \ \vdots \ \operatorname{commute\ time}_n \end{bmatrix}$$

• Then, all we need to do is solve the **normal equations**:

$$oldsymbol{X^TX}ec{w}^* = oldsymbol{X^Tec{y}}$$

If  $X^TX$  is invertible, we know the solution is:

$$ec{w}^* = (X^TX)^{-1}X^Tec{y}$$

# X1 - 1st date point EIR2

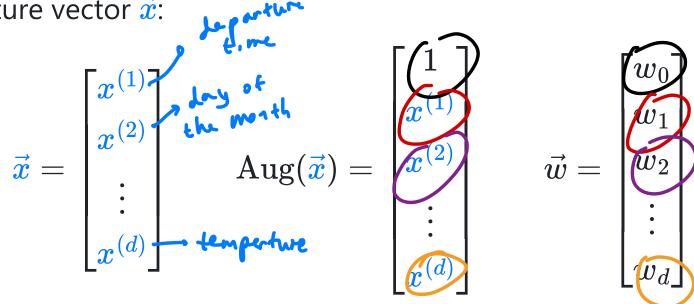
#### Notation for multiple linear regression

- X(1) 1st feature EIR"
- We will need to keep track of multiple features for every individual in our dataset.
  - In practice, we could have hundreds or thousands of features!
- ullet As before, subscripts distinguish between individuals in our dataset. We have nindividuals, also called training examples.
- Superscripts distinguish between **features**. We have d features.  $X^{(i)}$  ,  $X^{(i)}$  ,  $X^{(i)}$ departure hour:  $x^{(1)} \in \mathbb{R}^n$

day of month:  $x^{(2)} \in \mathbb{R}^{^{*}}$ 

#### Augmented feature vectors

• The augmented feature vector  $\operatorname{Aug}(\vec{x})$  is the vector obtained by adding a 1 to the front of feature vector  $\vec{x}$ :



• Then, our hypothesis function is:

$$H(\vec{x}) = w_0 + w_1 x^{(1)} + w_2 x^{(2)} + \dots + w_d x^{(d)}$$
 $= \vec{w} \cdot \operatorname{Aug}(\vec{x})$ 
 $= \vec{w} \cdot \operatorname{Aug}(\vec{x})$ 

#### The general problem

• We have n data points,  $(\vec{x}_1, y_1), (\vec{x}_2, y_2), \ldots, (\vec{x}_n, y_n)$ , where each  $\vec{x}_i$  is a feature vector of d features:

$$ec{x}_i = egin{bmatrix} x_i^{(1)} \ x_i^{(2)} \ dots \ x_i^{(d)} \end{bmatrix}$$
 EV

SLR model

(X1, y1), ... (X1, yn)

Xielh
scalar

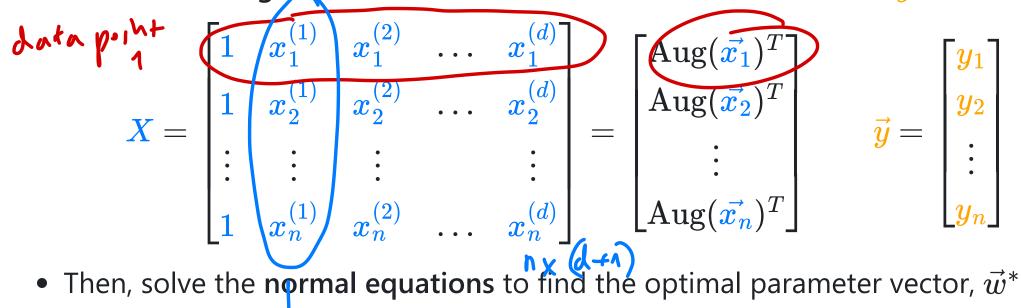
• We want to find a good linear hypothesis function:

$$H(\vec{x}) = w_0 + w_1 x^{(1)} + w_2 x^{(2)} + \ldots + w_d x^{(d)}$$

$$= \vec{w} \cdot \operatorname{Aug}(\vec{x})$$
How to find  $\vec{w} = (w_0, w_1, \ldots, w_d)$ 

#### The general solution

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



• Then, solve the **normal equations** to find the optimal parameter vector,  $\vec{w}^*$ :

$$X^T X \vec{w}^* = X^T \vec{y}$$
Feature
If  $X^T X \vec{w}^* = X^T \vec{y}$ 
optimal  $Y^2 = (X^T X)^{-1} X^T \vec{y}^*$ 

#### Terminology for parameters

- With d features,  $\vec{w}$  has d+1 entries.
- $w_0$  is the bias, also known as the intercept.
- $w_1, w_2, \ldots, w_d$  each give the **weight**, or **coefficient**, or **slope**, of a feature.

$$H(\vec{x}) = w_0 + w_1 x^{(1)} + w_2 x^{(2)} + \ldots + w_d x^{(d)}$$

### Interpreting parameters

#### **Example: Predicting sales**

- For each of **to** stores, we have:
  - net sales,square feet,
  - o inventory,
  - advertising expenditure,
  - district size, and
  - o number of competing stores.
- Goal: Predict net sales given the other five features.
- To begin, we'll start trying to fit the hypothesis function to predict sales:

 $H(\text{square feet, competitors}) = w_0 + w_1 \cdot \text{square feet} + w_2 \cdot \text{competitors}$ 

27×6

#### Question 👺

#### Answer at q.dsc40a.com

 $H(\text{square feet, competitors}) = w_0 + w_1 \cdot \text{square feet} + w_2 \cdot \text{competitors}$ 

What will be the signs of  $w_1^*$  and  $w_2^*$ ?

• A. 
$$w_1^* + w_2^* +$$

$$oldsymbol{lack}ullet$$
 B.  $w_1^*+ \qquad w_2^*- \ oldsymbol{lack}ullet$  A.  $w_1^*- \qquad w_2^*+ \ oldsymbol{lack}$ 

$$ullet$$
 A.  $w_1^st - w_2^st +$ 

$$ullet$$
 A.  $w_1^*$   $w_2^*$   $-$ 

Let's find out! Follow along in this notebook.