Linear Regression

$x \longrightarrow y$

x is a vector

$x \in \mathbb{R}^d \rightarrow x$ is a vector of length d

y is a scalar

$y \in \mathbb{R}$

Prediction or Regression

Predict future income

Current GPA, # AI courses taken	y
	$y = 7000$
	$y = 300000$

3.8	4

y ?

Training data

GPA, # AI	Income
	100000

3.7, 2	100000
3.9, 6	500000
2.4, 1	1,000,000

3.5, 2	?

Functional models.

$$y = f(\mathbf{x})$$

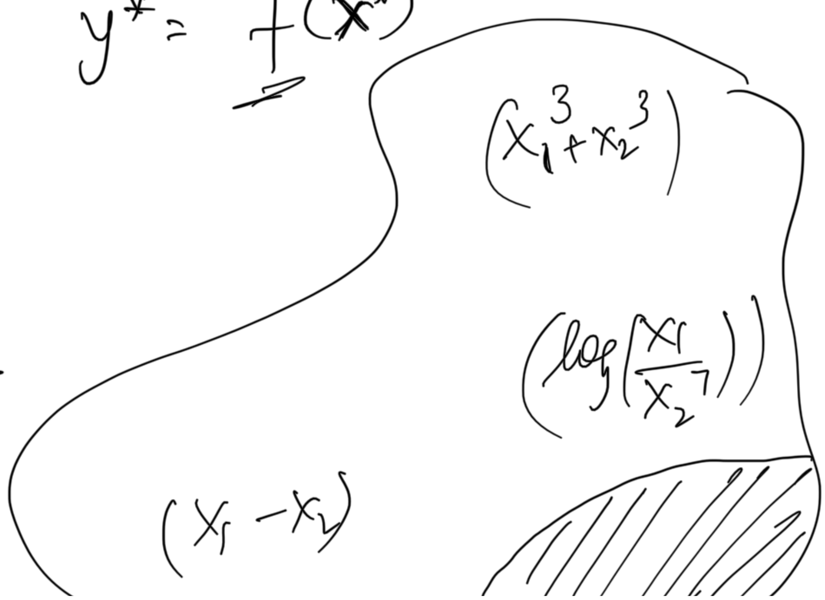
some function.

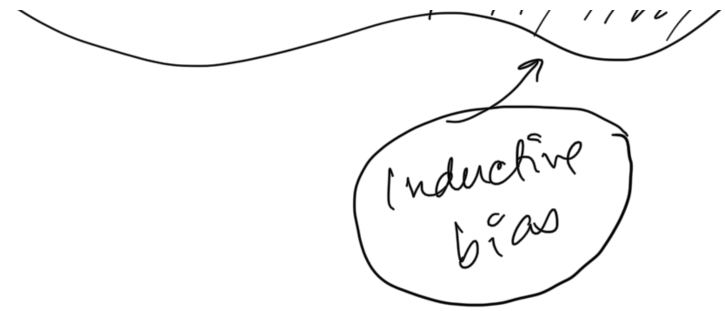
If we learn $f(\,)$,

then for a new $\mathbf{x}^*$ $\quad \mathbf{x} = \boxed{x_1, x_2}$

$$y^* = \underline{f}(\mathbf{x}^*)$$

a big bag of functions





Monday Feb 8

$$\mathbf{X} \longrightarrow Y$$

Functional models

$$Y = \underset{\sim}{f}(\mathbf{X})$$

Probabilistic Models.

$$p(\mathbf{X}, Y)$$

$$p(Y \mid \mathbf{X}) \leftarrow \text{Bayes Rule}$$

X_1

$$\mathbf{X} \rightarrow \boxed{x_1 \mid x_2 \mid \ \mid \ \mid \ \mid \ \mid \ \cdots x_D}$$

$\mathbf{X}, \mathbf{Y}, \mathbf{Z}$

$\mathbf{X}_1 \quad \mathbf{X}_2$ $\qquad \boxed{X_{11} \quad X_{12} \ \cdots}$

$$\sum^{D} x_i y_i$$

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1} x_i y_i$$

$$= x_1 y_1 + x_2 y_2 + \cdots + x_D y_D$$

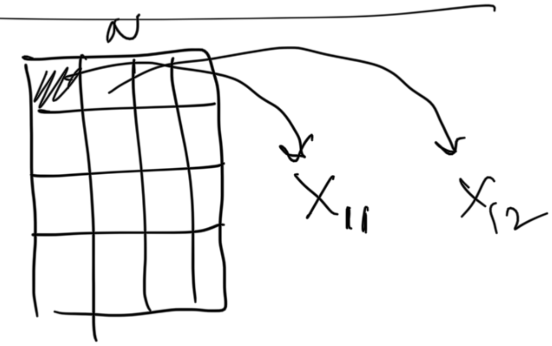
$$|\mathbf{x}| = \sum_{i=1}^{D} |x_i|$$

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{D} x_i^2} \quad l_2 \text{ norm}$$

$\|\mathbf{x}\|$

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{D} x_i^{\,p} \right)^{1/p}$$

<u>Matrix</u>

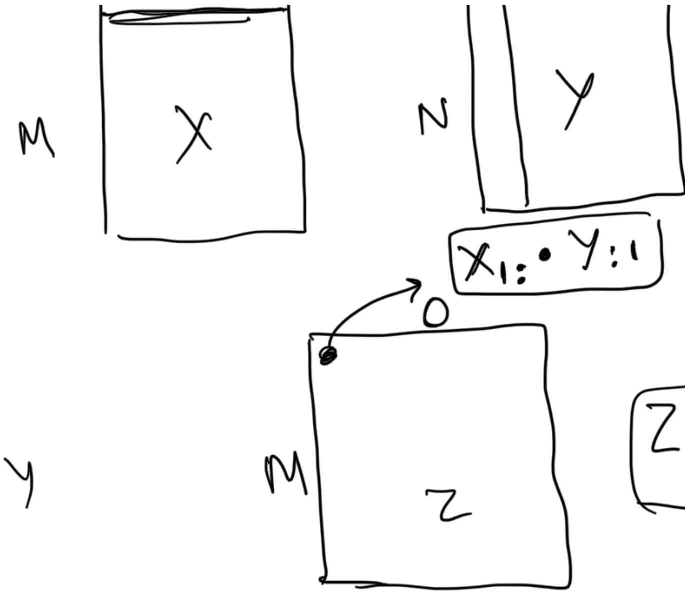


$$\boxed{\mathbf{X} \in \mathbb{R}^{M \times N}}$$

$$Y = X^T \quad (N \times M)$$

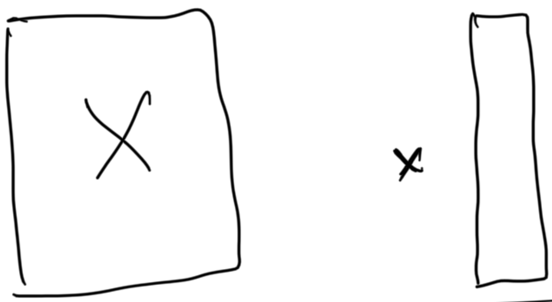
$$c\mathbf{X} = \begin{bmatrix} cx_{11} & cx_{12} & \cdots \\ \vdots & & \vdots \\ \vdots & & \vdots \end{bmatrix}$$

$\mathbf{X}Y$



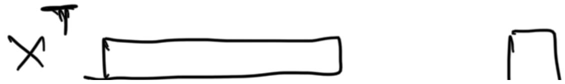
$Z = XY$

$$\boxed{Z_{ij} = \underline{X_{i:} \cdot Y_{:j}}}$$



A D-length vector is a $(D \times 1)$ matrix

$$x \cdot y = x^T y$$

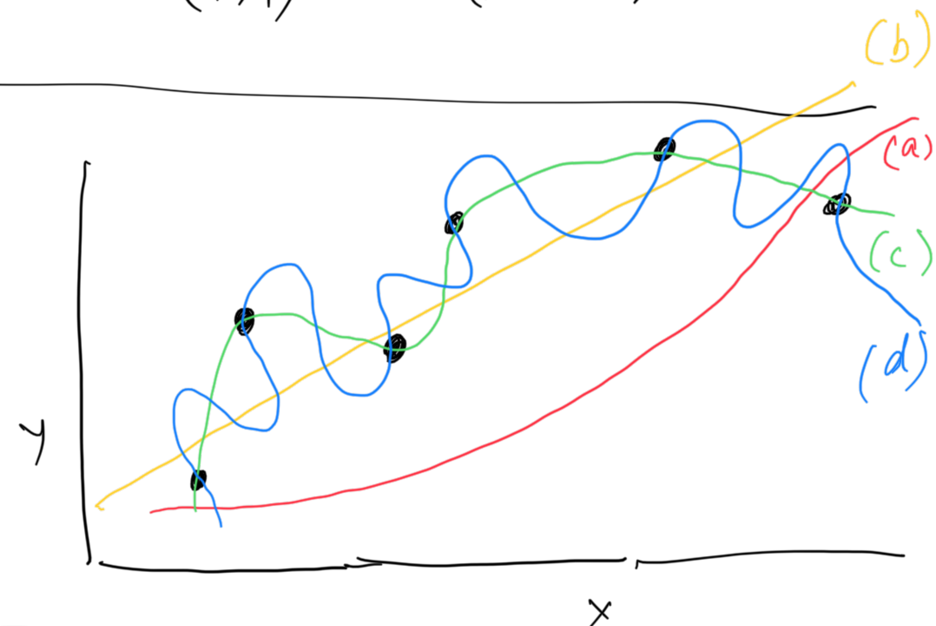


$1 \times D$

y

$D \times 1$

$$z = x^T y = (1 \times 1) = (x^T) \cdot y$$



error

$$(b) \rightarrow \quad y = mx + c$$

$$(c) \rightarrow \quad y = ax^3 + bx^2 + cx + d$$

$$(d) \rightarrow \quad y = ax^{10} + bx^9 + \cdots + \text{---}$$

Linear Regression

$y = f(\mathbf{x})$

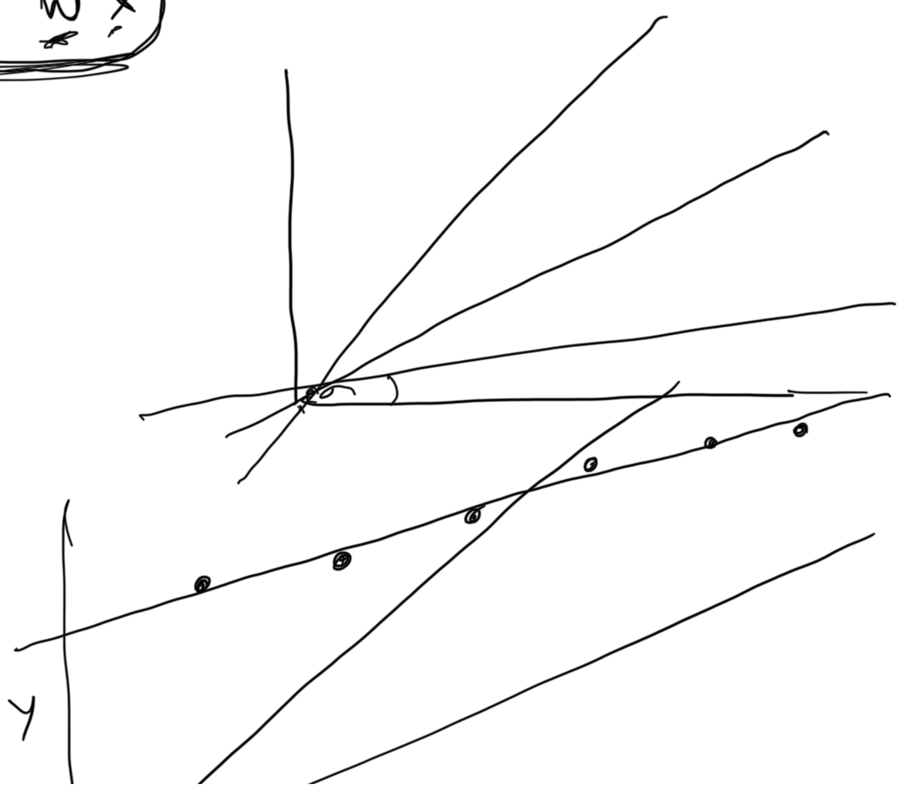
$$y = \underline{\mathbf{w}^T\mathbf{x}} \rightarrow D \times 1$$

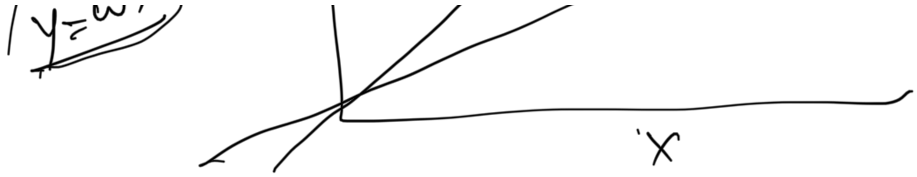
weight vector $(D \times 1)$

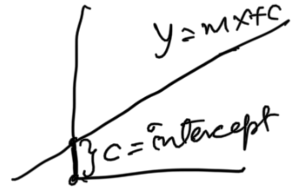
$$= w_1 x_1 + w_2 x_2 + \cdots + w_D x_D$$

For the flu example

$$\boxed{y = w x}$$







$$\boxed{y = w_0 + w_1 x}$$

bias-term

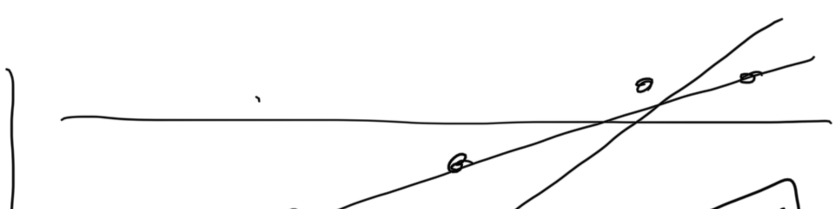
Given some data:

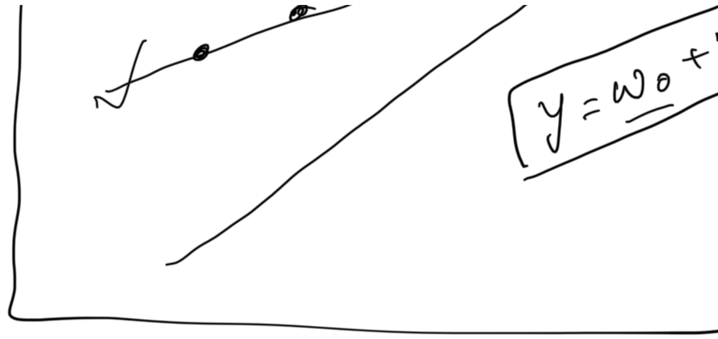
x	y
12	9
28	15
15	11
48	21
56	22

Find the best w_0, w_1 which "fits" the data best.

Wed, Feb 10

Resources on Piazza





$\boxed{y = \underline{w_0} + \underline{w} x}$

Find w_0, w that does <u>what</u>?

X	Y	Pred. $\bar{y}$
x_1	y_1	$\bar{y}_1$
x_2	y_2	$\bar{y}_2$
x_3	y_3	$\bar{y}_3$
⋮	⋮	⋮
x_N	y_N	$\bar{y}_N$

For a given w, w_0

$$\bar{y}_1 = w_0 + w x_1$$

⋮

$$\bar{y}_i = w_0 + w x_i$$

Error: $e_i = y_i - \bar{y}_i$

$$J = \frac{1}{\ } \sum^{N} e_i^2$$

$2 \quad i=1$

$$\boxed{J(w_0, w) = \frac{1}{2} \sum_{i=1}^{N} \left(y_i - (w_0 + w x_i)\right)^2}$$

Squared loss function

$\frac{1}{2} \rightarrow$ just for mathematical convenience

---

what if x is not 1-D

$\Rightarrow \quad x \in \mathbb{R}^d$ where $d > 1$

$y = w_0 + \mathbf{w}^T x$

$\mathbf{w} \in \mathbb{R}^d$

Squared loss function

$$\boxed{J(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \mathbf{w}^T x_i\right)^2}$$

absorbed

added a 1 to x_i

w_0 into w

Find w that minimizes $J(w)$

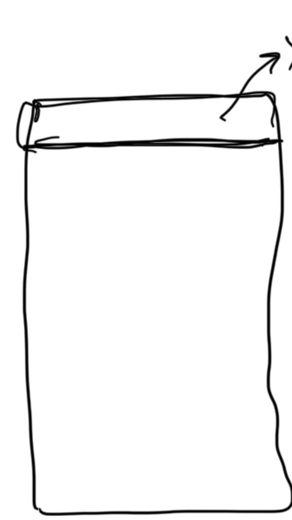
$$w^T x_i = \sum_{j=0}^{d} (w_j \, x_{ij})$$

$w \cdot x_i$

$w^T x_i$

x_1	y_1
x_2	y_2
x_3	y_3
$\vdots$	$\vdots$
x_N	y_N

y is a vector ($N \times 1$)



X is a matrix ($N \times (d+1)$)

X

Training data: X, y — target vector

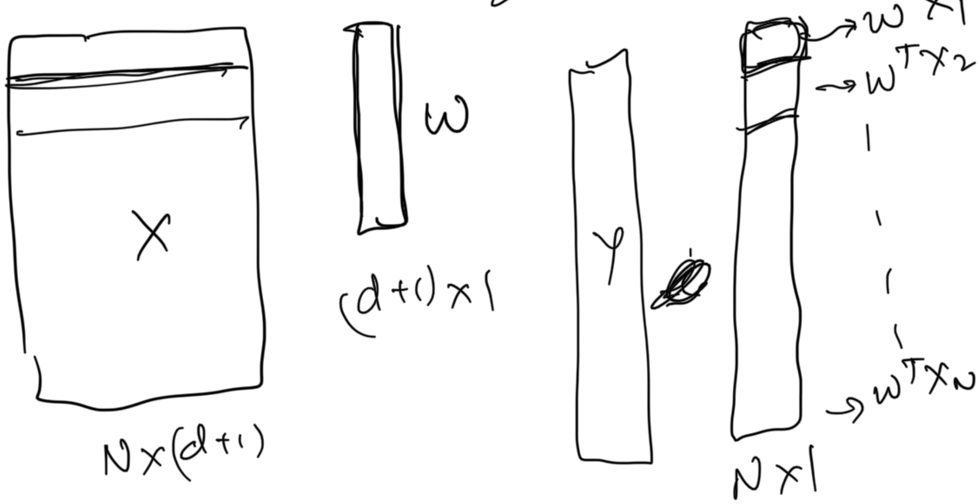
($N \times (d+1)$) ($N \times 1$)

data matrix

$$J(w) = \frac{1}{2} \sum_{i=1}^{N} (y_i - w^T x_i)^2$$

$$\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_N \end{array} \quad \begin{array}{c} y_1 - w^T x_1 \\ y_2 - w^T x_2 \\ \vdots \\ y_N - w^T x_N \end{array} \qquad e = \begin{bmatrix} y_1 - w^T x_1 \\ y_2 - w^T x_2 \\ \vdots \\ y_N - w^T x_N \end{bmatrix}$$

$$e = (Y - Xw)$$



$$\underbrace{\sum (y_i - w^T x_i)^2}_{e} = (Y - Xw)^T (Y - Xw)$$

$$e_i = y_i - w^T x_i$$

$$\mathbf{e} = \begin{bmatrix} y_1 - w^T x_1 \\ \vdots \\ y_N - w^T x_N \end{bmatrix}$$

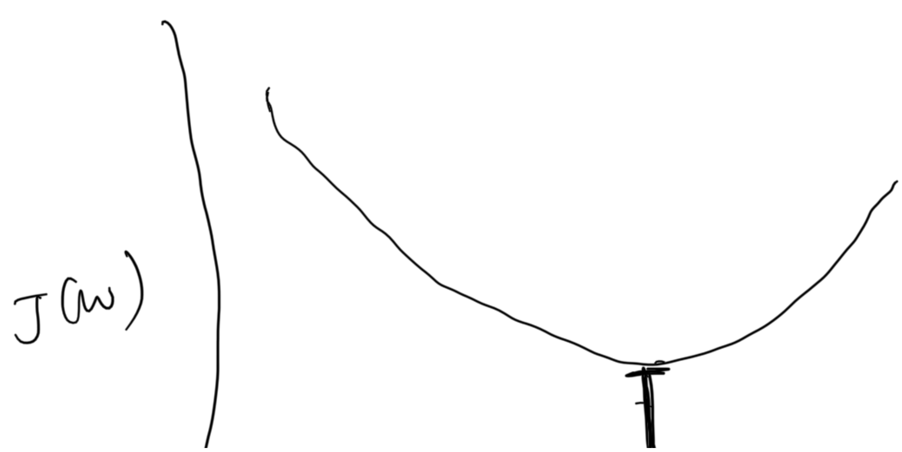
any vector $\mathbf{p}$

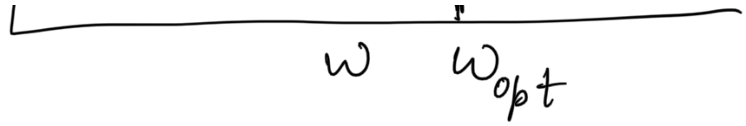
$$p^T p = \sum p_i^2$$

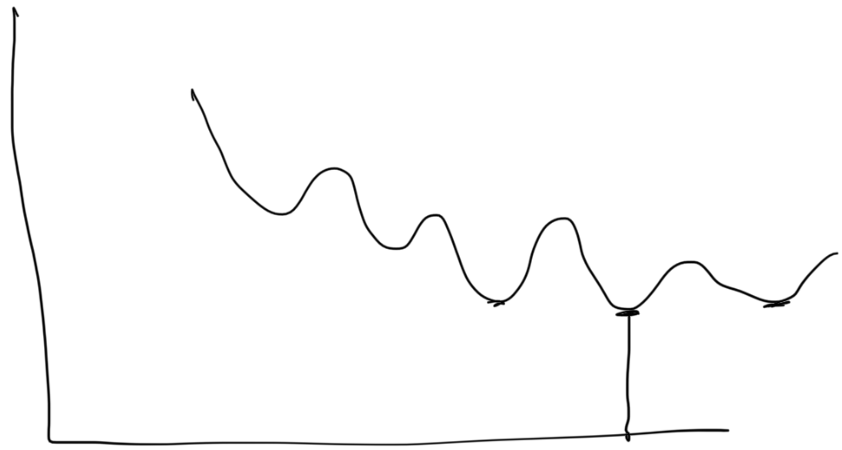
$$\sum (y_i - w^T x_i)^2 = (\mathbf{y} - \mathbf{X}\mathbf{w})^T (\mathbf{y} - \mathbf{X}\mathbf{w})$$

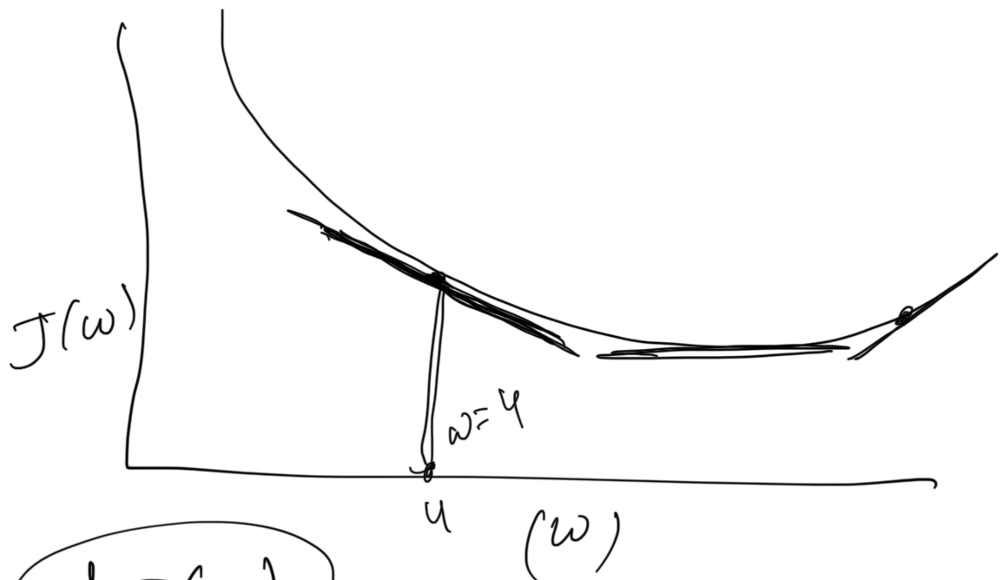
$$\boxed{J(w) = \frac{1}{2}(y - Xw)^T (y - Xw)}$$

How do we find $\mathbf{w}$ that minimizes $J(\mathbf{w})$









$$\frac{d}{dw} J(w)$$

e.g. $J(w) = 3w^3 + 4w$

$$\frac{dJ(w)}{dw} = 9w^2 + 4$$

at $w = 4$

$$\frac{dJ(w)}{dw} = \underline{148}$$

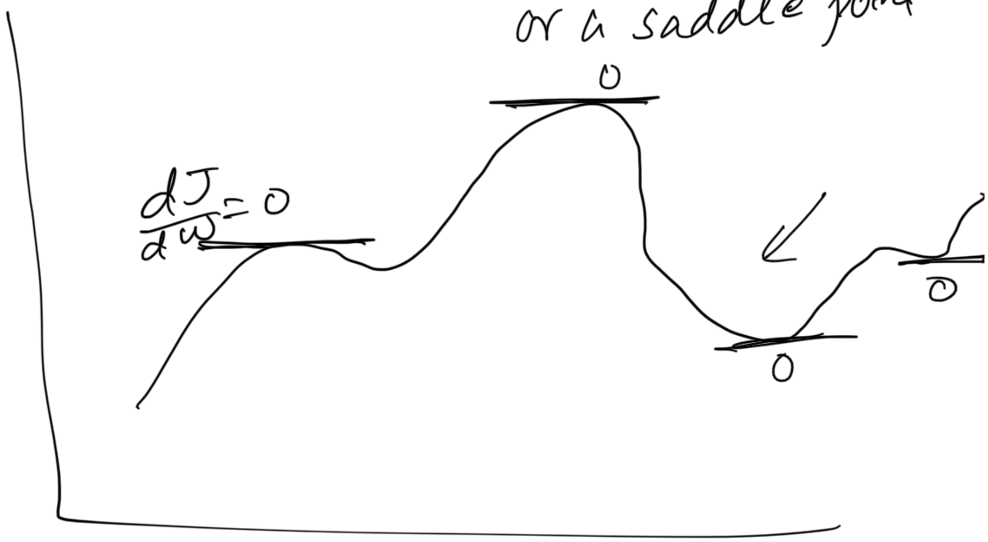
If d J(w)

If $\frac{d\,J(w)}{dw} = 0$ at a given w

that means $w \to$ point of minima

or maxima

or a saddle point



If $J(w)$ is convex

then $\boxed{\frac{d}{dw} J(w) = 0}$

solution for this will give us w for which $J(w)$ is minima