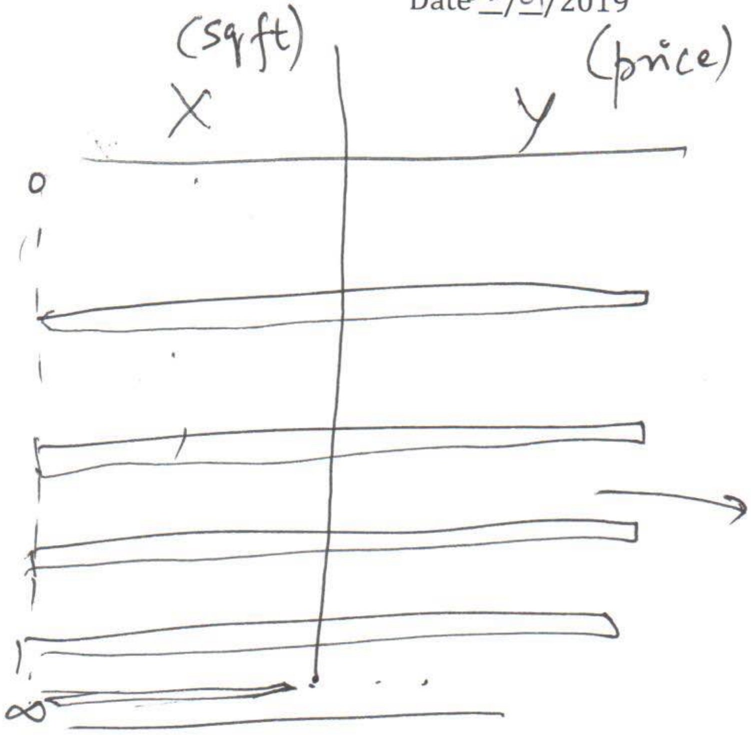
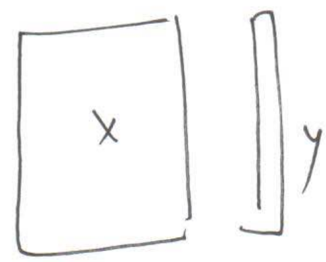


$$y = \underline{f}(x)$$

Training data.



$$x^* \xrightarrow[f()]{} y^*$$

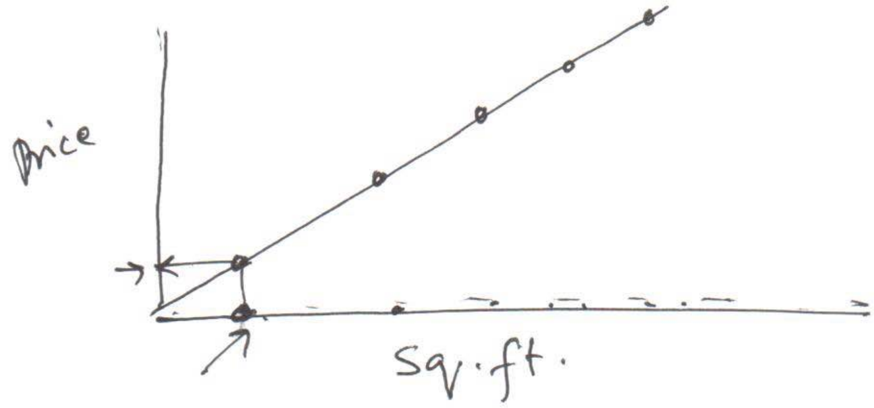
Linear Regression

$f()$ is a linear fn of x.

Inductive Bias

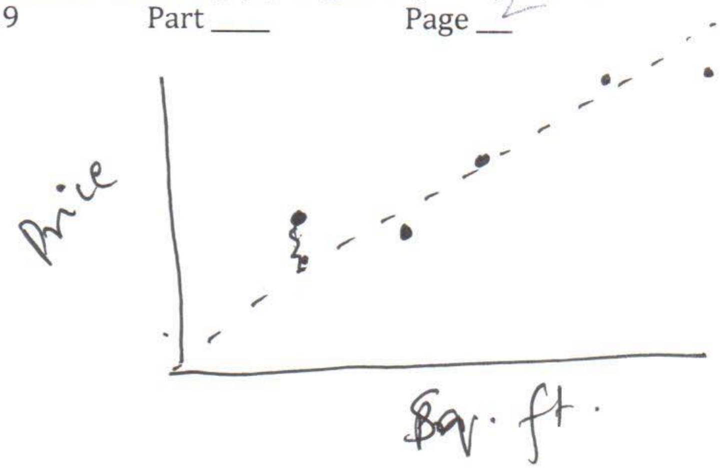
$$\boxed{y = wx}$$

w is a scalar value.



Loss function

↳ Optimization



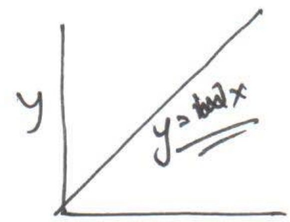
Training Data.

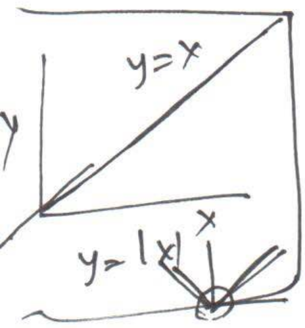
X sq.ft.	Y price (1000 $)
100	500
300	800
200	700
400	950

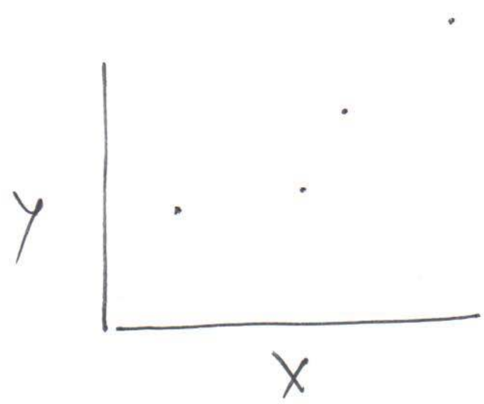
$y = f(x)$

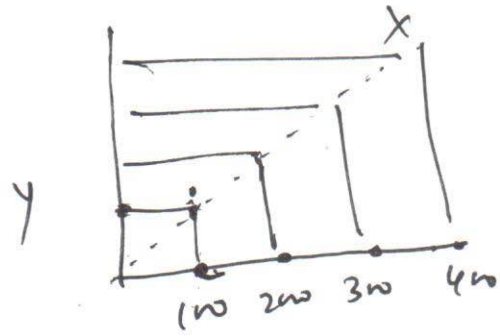
Inductive Bias.

$y = w\,x$







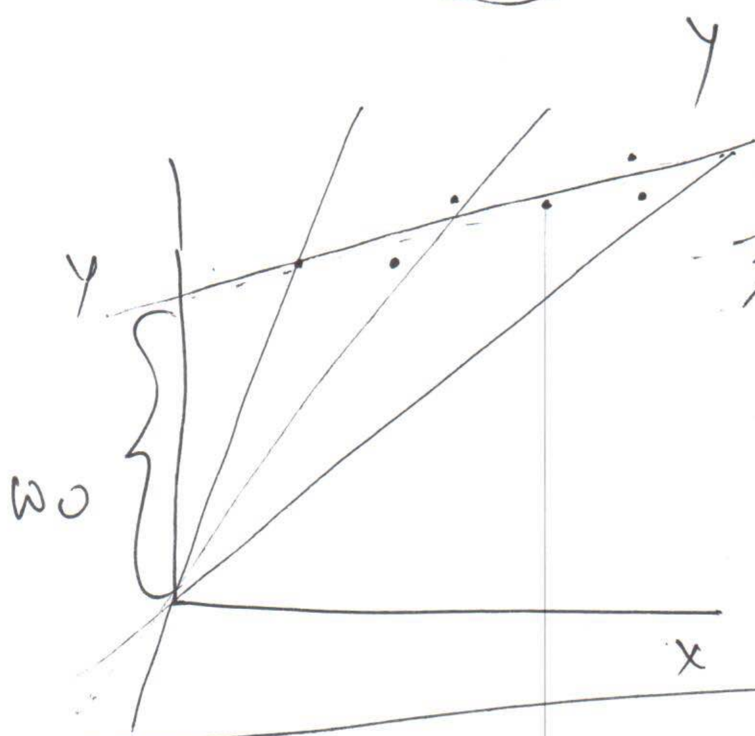


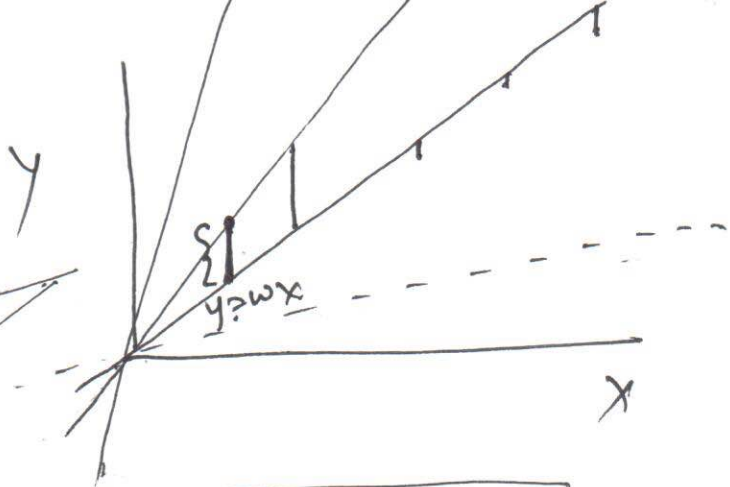
let $w = 4$

	X	Y		$\hat{y}^* = wX$		
x_1	100	500	y_1	400	300	$(y_1 - wx_1)^2$ +
x_2	300	800	y_2	1200	900	$(y_2 - wx_2)^2$ +
x_3	200	700	y_3	800	600	$(y_3 - wx_3)^2$ +
x_4	400	950	y_4	1600	1200	$(y_4 - wx_4)^2$
				$w=4$	$w=3$	↑ squared loss

loss fn $\boxed{J(w) = \frac{1}{2}\sum_{i=1}^{n} (y_i - w x_i)^2}$

$y = wx$





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

$y = w_0 + w_1 x$

Intercept. → Bias term

$w_0 = 200,\ w_1 = 4$

x	y	
100	500	~~200+4×~~ 600
300	800	1400
200	700	1000
400	950	1800

$$\boxed{J(w) = \frac{1}{2} \sum_{i=1}^{n} \left[\left(y_i - (w_0 + w_1 x_i) \right)^2 \right]}$$

X	Y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_n	y_n

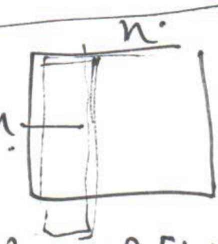
$$d \begin{bmatrix} 0.04 \\ 3 \\ 45 \\ \vdots \end{bmatrix}$$

$\underline{x \in \mathbb{R}^d}$

$$\boxed{y_i = w_0 \cdot 1 + w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id}}$$

$$W \to \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}$$

$\underset{m \times n}{A} =$



$A^T = B \qquad B[i,j] = A[j,i]$

W [column] $\qquad W^T =$ [row]

$\underline{A = A^T} \to A$-symmetric

$\begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \end{bmatrix} \to$ ~~Diagonal~~ Identity Matrix

$A\,B \to$ Matrix Multiplication.

Matrix Algebra Refresher.

$W \quad V \quad \in \mathbb{R}^d$

vectors.

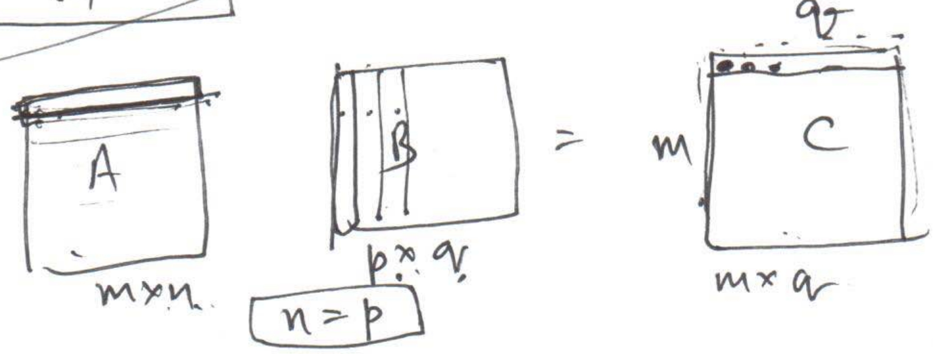
$$W \cdot V = \sum_{i=1}^{d} w_i v_i$$

dotproduct or inner product.

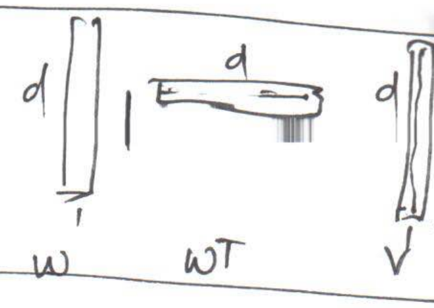
$\underline{W^T V} = W \cdot V$

$$\boxed{W \cdot W = \sum_{i=1}^{d} w_i w_i = \sum_{i=1}^{d} w_i^2}$$

Matrix Mult. :



$$w \cdot v = \underbrace{w^T v}_{\text{Matrix Multiplication}}$$



$$y_i = \underline{w_0} + w_1 x_{i1} + w_2 x_{i2} + - - - + w_d x_{id}$$

$$y_i = w_0 + \mathbf{w}^T \underline{\mathbf{x}_i} \qquad \mathbf{w} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}$$

let $$x_i \equiv \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \end{bmatrix}$$

$$\boxed{y_i = \mathbf{W}^T \mathbf{x}_i}$$ where x_i and $$\mathbf{W} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}$$

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

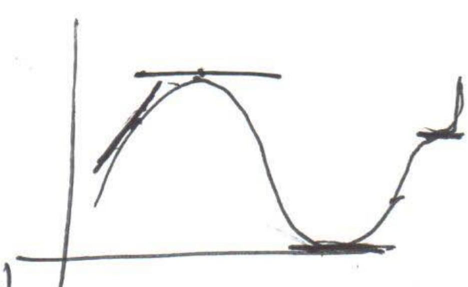
learn w that minimizes J(w)

finding function extremes.

$$f(x) = \underline{x^3 - 2x}$$

$$f'(x) = \frac{d}{dx} f(x) = 0$$

$$f''(x) = \frac{d}{dx}\frac{d}{dx} f(x) = \underline{\frac{d^2}{dx^2} f(x)}$$

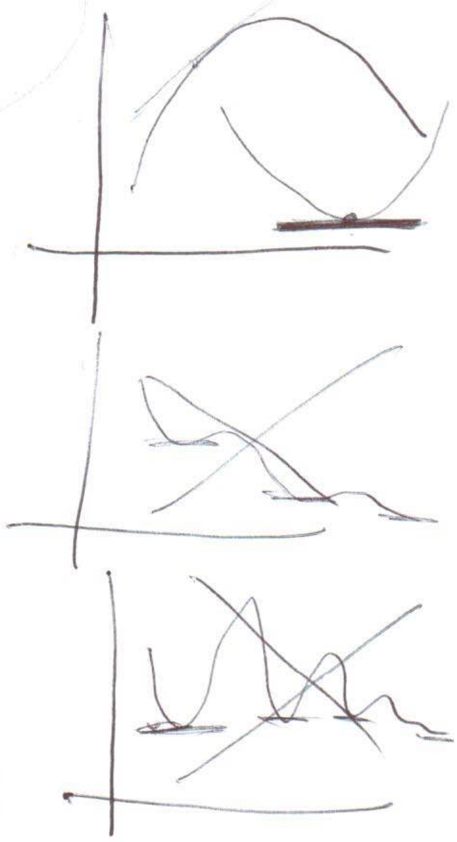


bias term	Sq.f	Price
1	~~400~~	200
1	300	100
1	900	500
1	850	400

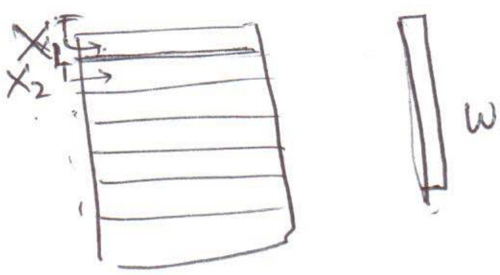
$\underset{4\times 2}{X} \qquad \underset{4\times 1}{y}$

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

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



$$\underset{n\times d}{X}\,\underset{d\times 1}{w} = n\times 1$$



$$\begin{bmatrix} x_1^T w \\ x_2^T w \\ x_3^T w \\ \vdots \end{bmatrix} = \begin{bmatrix} w^T x_1 \\ w^T x_2 \\ \vdots \end{bmatrix}$$

$a^T b = b^T a$

$$y - Xw = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix} - \begin{bmatrix} w^T x_1 \\ \vdots \end{bmatrix}$$

$a^T a = \sum a_i^2$

$$= \begin{bmatrix} y_1 - w^T x_1 \\ y_2 - w^T x_2 \\ \vdots \end{bmatrix}$$

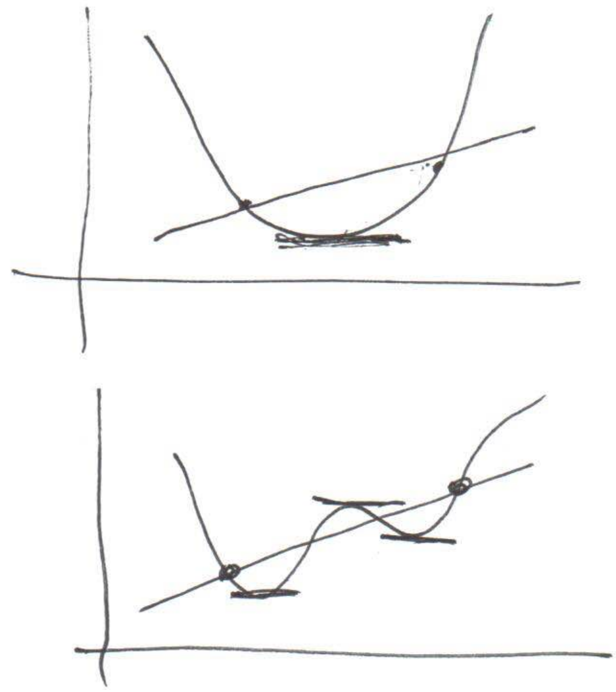
$$(y - Xw)^T(y - Xw) = \frac{1}{2}\sum_{i=1}^{n}(y_i - w^T x_i)^2$$

$J(w) \rightarrow$ Smooth

$\rightarrow$ Continuous.

$\rightarrow$ Convex

Lips'- - -



Gradient

$$f(x,y) = 7x^3 + 2xy + 4x$$

$$\frac{\partial}{\partial x} f(x,y) = 21x^2 + 2y + 4$$

$$\frac{\partial}{\partial y} f(x,y) = 2x$$

$$\nabla f = \begin{bmatrix} \frac{\partial}{\partial x} f(x,y) \\ \frac{\partial}{\partial y} f(x,y) \end{bmatrix} = \begin{bmatrix} 21x^2+2y+4 \\ 2x \end{bmatrix}$$

$$\underline{\nabla f = 0}$$

$$\begin{bmatrix} 21x^2+2y+4 \\ 2x \end{bmatrix} = 0$$

$$21x^2 + 2y + 4 = 0$$

$$2x = 0$$

$$\boxed{\begin{aligned} f(x) &= 7x^2 + 4x \\ \frac{d}{dx} f &= 14x + 4 \\ 14x + 4 &= 0 \\ x &= \frac{-4}{14} \end{aligned}}$$

$$f(w_0, w_1) = (a - w_0 + w_1)^2$$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial w_0} \\ \frac{\partial f}{\partial w_1} \end{bmatrix}$$

$$\frac{\partial f}{\partial w_0} = \frac{\partial}{\partial w_0} (a - w_0 + w_1)^2$$

$$= 2(a - w_0 + w_1)(-1)$$

$$\frac{\partial f}{\partial w_1} = \frac{\partial}{\partial w_1} (a - w_0 + w_1)^2 = 2(a - w_0 + w_1).$$

$$\frac{d}{dw}\frac{1}{2}\left[(y - Xw)^T (y - Xw)\right]$$

$$\frac{1}{2}\frac{d}{dw}\left((y^T - (Xw)^T)(y - Xw)\right)$$

$$= \frac{1}{2}\frac{d}{dw}\left((y^T - \underline{w^T X^T})(y - \underline{Xw})\right)$$

same

$$= \frac{1}{2}\frac{d}{dw}\left[\cancel{y^T y} + w^T X^T X w - \underbrace{w^T X^T y}_{1\times d \,\times\, d\times n \,\times\, n\times 1} - y^T X w\right]$$

$$= \frac{1}{2}\left[\frac{d}{dw}(w^T X^T X w) - \frac{d}{dw}(2 w^T X^T y)\right]$$

$$= \frac{1}{\cancel{2}}\left[\cancel{2} X^T X w - \cancel{2} X^T y\right]$$

$$= X^T X w - X^T y$$

Setting gradient of $J(w)$ to 0

$$X^T X w - X^T y = 0$$

$$X^T X w = X^T y$$

$$(X^T X)^{-1}(X^T X w) = (X^T X)^{-1} X^T y$$

$$\boxed{w = (X^T X)^{-1} X^T y}$$

X - $n \times d$

X^T - $d \times n$

$X^T X = d \times d$

symmetric

$$\frac{d}{da} a^T M a = 2Ma$$

$$\frac{d}{da} a^T b = b$$

$$\boxed{X^{-1} X = I}$$

x

$y = \frac{1}{x}$ $x \to 0$

$x = 0.00000000\ldots 1$

$y = \frac{1}{x}$

Z

Z^{-1} np.linalg.inv(Z)

$Z \to$ Singular

$$\begin{bmatrix} 4 & 3 & 1 \\ 4 & 3 & 7 \\ 4 & 3 & 9 \\ 4 & 3 & 1 \end{bmatrix}$$

$Z^{-1} y = a$

$Z \begin{bmatrix} a \end{bmatrix} = y$

(np.linalg.inv(Z) * y)

np.linalg.solve(Z, y)

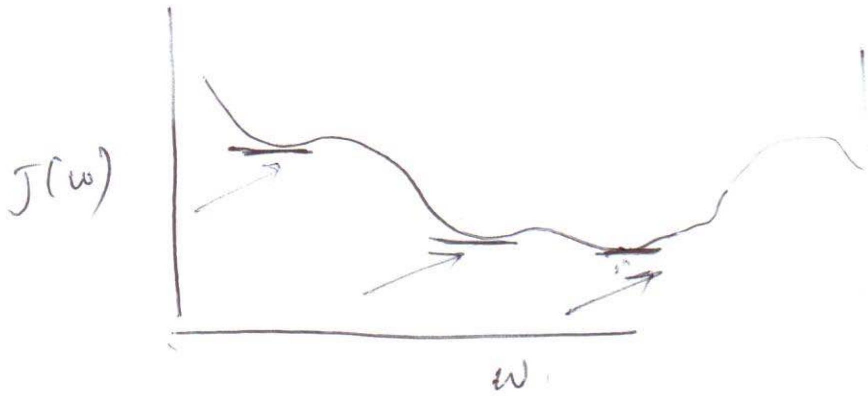
$\boxed{\text{Gradient Descent}}$

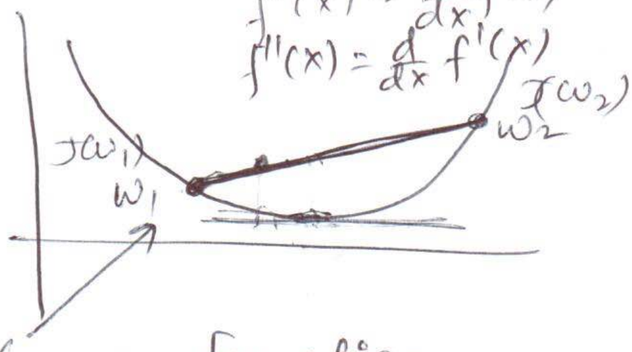
Friday

$f(x)$

$f'(x) = \frac{d}{dx} f(x)$

$f''(x) = \frac{d}{dx} f'(x)$





Convex functions.

$$J(w) = (y - Xw)^T (y - Xw)$$

$$\nabla_w J = X^T X w - X^T y$$

$$\frac{d}{dw} \nabla J = \begin{bmatrix} \quad \end{bmatrix} = H$$

positive definite

Hessian

$$w = (X^T X)^{-1} X^T y$$

X - $n \times d$ $\qquad X^T X \to d \times d$ matrix

X^T - $d \times n$

Matrix inversion - $O(d^3)$ → for a $d \times d$ matrix.

Gradient Descent Based Linear Regression.

$$J(w) = (y - Xw)^T (y - Xw)$$

$$\nabla J = \underbrace{\boxed{X^T X}}_{d \times d} w - X^T y \longrightarrow O(d^2)$$

For k steps: $\cancel{k\,O}$ $O(kd^2)$

$O(nd)$.

$\cancel{X^T}$ $\quad X^T X \rightarrow O(nd)$

$d \times d \quad \times d \times 1 \rightarrow O(d^2)$

Stochastic Gradient Descent.

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

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

initialize w

for $i = 1$ to n :

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

Gradient Descent for $J(w_i)$