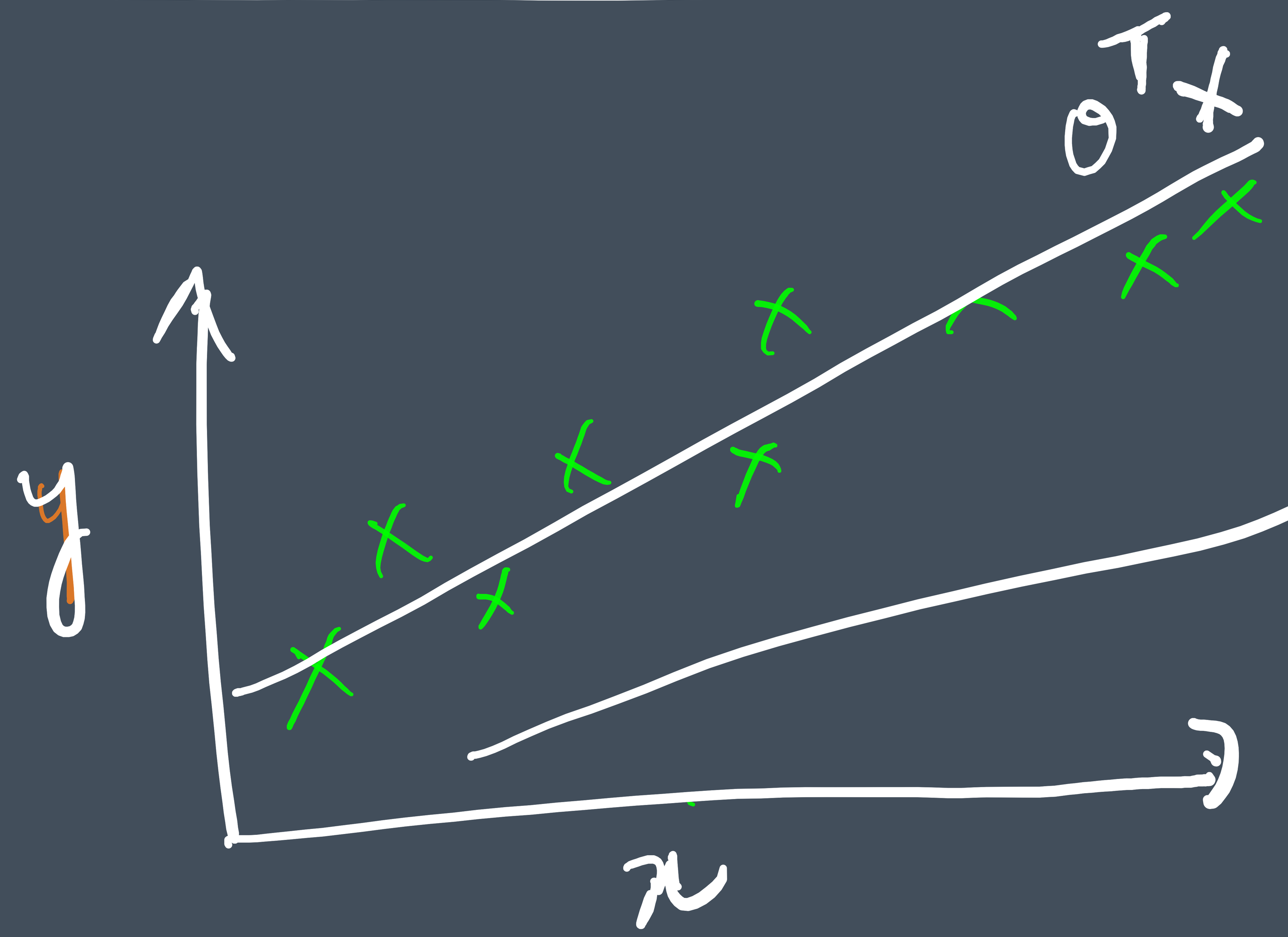
Regression

Closed Form Solution



Best Fit

$\theta = [\theta_0 \ \theta_1]$

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

$y = \theta_1 x + \theta_0$

m : # examples

n : # no of features

$$X = \begin{bmatrix} \text{---} \ x^{(1)} \ \text{---} & 1 \\ \text{---} \ x^{(2)} \ \text{---} & 1 \\ \text{---} \ x^{(i)} \ \text{---} & 1 \\ \vdots & 1 \\ \text{---} \ x^{(m)} \ \text{---} & 1 \end{bmatrix}$$

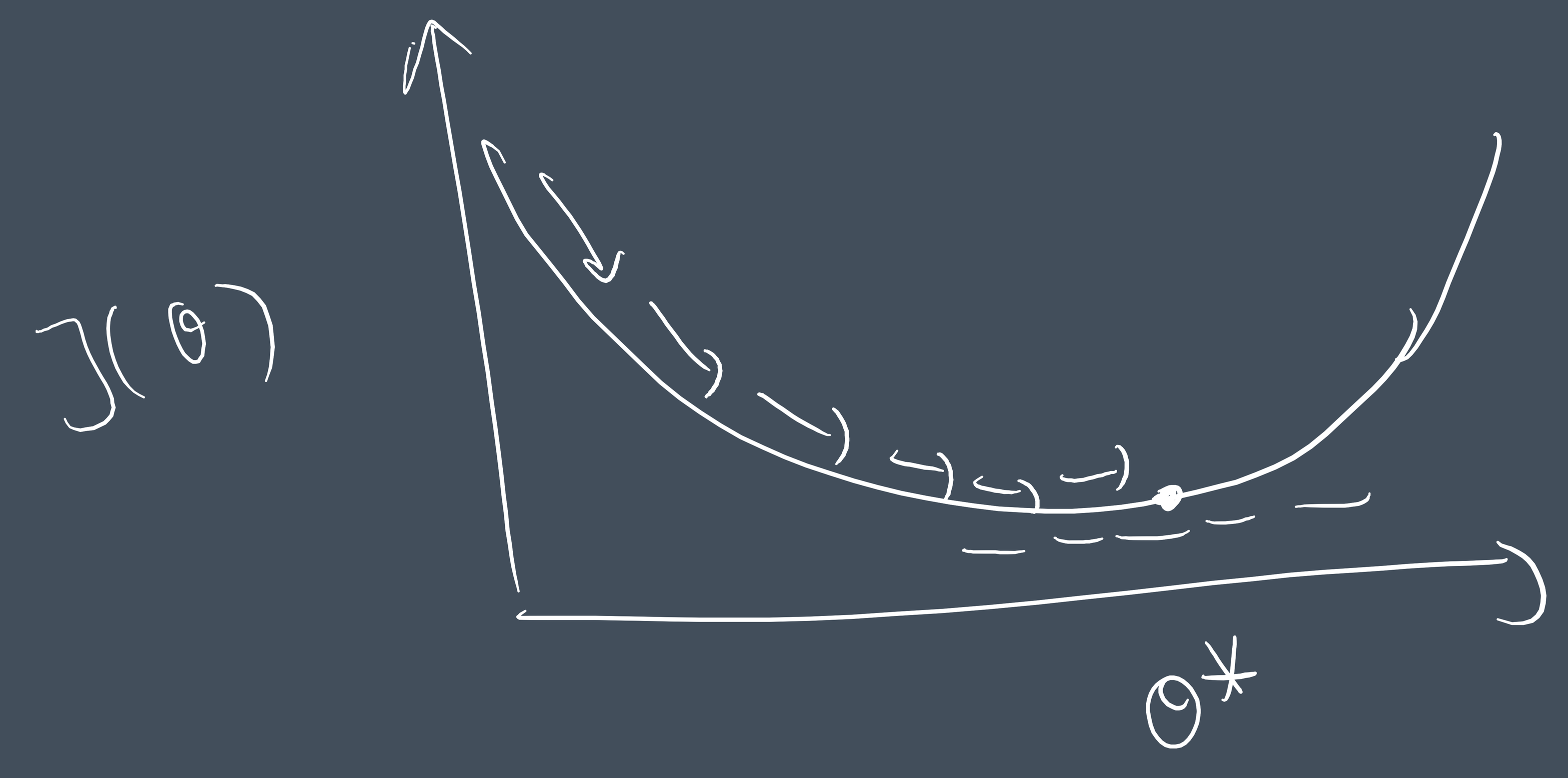
$m \times (\underline{n} + 1)$

$x_j^{(i)}$ = i^{th} example, j^{th} feature.

$$\theta = \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_n \end{bmatrix}_{(n+1)}$$

Proof of Closed Form of Regression

Gradient Desent



$$\text{Loss} = \frac{1}{2}\sum_{i=1}^{m}\left(y^{i}_{pred} - y^{i}\right)^2$$

$$y_{pred} = h_\theta(x^{i}) = \sum_{i=0}^{n}\theta_i x_i^{j}$$

$$X = \begin{array}{c} x_0 \\ \left[\begin{array}{c} \text{---}\; x^{(1)} \;\text{---} \\ \vdots \\ \text{---}\; x^{(i)} \;\text{---} \\ \vdots \\ \text{---}\; x^{(m)} \;\text{---} \end{array}\right]_{m\times(n+1)} \end{array} \qquad Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}_{m\times 1}$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}_{(n+1)\times 1}$$

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots\ldots$$

$$\boxed{x_0 = 1}$$

$$= \theta^T x = \sum_{i=0}^{m}\theta_i x_i$$

$$=$$

Proof

$$\underbrace{\text{loss} = \frac{1}{2}(X\theta - Y)^2}_{\text{matrix notation}} \Leftrightarrow \frac{1}{2}\sum_{i=1}^{m}\left(y_{pred}^{(i)} - y^{(i)}\right)^2$$

$m \Rightarrow m$

$n \Rightarrow 2$

$$\overset{X}{\begin{bmatrix} 1 & x_1^1 & x_2^1 \\ 1 & x_1^2 & x_2^2 \\ \vdots & & \\ 1 & x_1^m & x_2^m \end{bmatrix}} \overset{\theta}{\begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}} = \begin{bmatrix} \theta_0 + \theta_1 x_1^1 + \theta_2 x_2^1 \\ \theta_0 + \theta_1 x_1^2 + \theta_2 x_2^2 \\ \vdots \end{bmatrix} = \begin{bmatrix} y_{pred}^1 \\ y_{pred}^2 \\ \vdots \\ y_{pred}^m \end{bmatrix}$$

$$\Big(\underbrace{X\theta}_{y_{pred}} - \underbrace{y}_{y_{actual}}\Big)^2$$

$$\left(\begin{bmatrix} \\ \\ \\ \end{bmatrix}_{m\times 1} - \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{m\times 1} \right)^2 \Rightarrow \underline{\underline{\text{Scalar}}}$$

Square

$$\left(\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}\right)^2 \qquad X^2 = X^T X$$

$$\Downarrow$$

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1^2 + 2^2 + 3^2 = \underline{\underline{\text{Scalar}}}$$

$$J(\theta) = (X\theta - y)^2 \qquad z^2 = z^T z$$

$$= (X\theta - y)^T (X\theta - y)$$

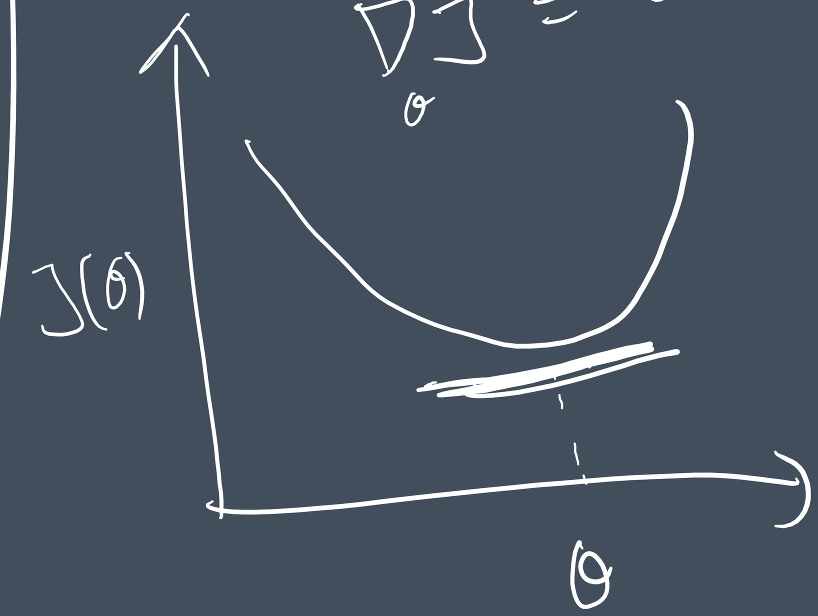
loss in matrix notation

$$(AB)^T = B^T A^T$$

$$= (\theta^T X^T - y^T)(X\theta - y)$$

$$J(\theta) = \underbrace{\theta^T X^T X \theta - \theta^T X^T y - y^T X \theta + y^T y}$$

$\nabla_\theta J = 0$



$$\nabla_\theta J(\theta) = 0$$

Compute first Derivative $\nabla_\theta J(\theta)$

$$\nabla_\theta J(\theta) = \nabla_\theta\left(\theta^T X^T X \theta - \theta^T X^T y - y^T X \theta + \cancel{y^T y}\right)$$

0

$$= 2X^T X \theta - \nabla_\theta\left(\theta^T X^T y + \theta^T X^T y\right)$$

$$= 2X^T X \theta - \nabla_\theta\left(2\theta^T X^T y\right)$$

$$\Rightarrow 2X^T X \theta - 2X^T y = 0$$

few Rules of matrix Calculus

$a \in \mathbb{R}^n$

1. $\nabla_\theta a^T \theta = a$
2. $\nabla_\theta \theta^T a = a$
3. $\nabla_\theta \theta^T A \theta = 2A\theta$

A is a $\mathbb{R}^{n \times n}$

$$\Rightarrow X^T \cdot X \theta = X^T y$$

$$\Rightarrow \underbrace{(X^T X)^{-1} (X^T X)}_{1} \theta) = (X^T X)^{-1} X^T y$$

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

⇓

closed form solution

$$\underbrace{y^T}_{1 \times m} \underbrace{X}_{m \times n} \underbrace{\theta}_{m \times 1} = 1 \times 1 \Rightarrow \text{scalar}$$

$$[5]^T = [5]$$

$$(y^T X \theta)^T = \underbrace{\theta^T X^T y}$$

$X = m \times n$

Gradient Descent

→ iterative, $\eta = ?$

→ Slow if n is very small.

→ Mini Batch

Closed form solution

→ Small dataset, directly get the solution.

→ expensive large dataset

→