Omar Ghattas Tutorials (Youtube)

Recap: Chain Rule

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

Ex: $y = (2x^2 + 3x)^4$

$$\frac{dy}{dx} = (4x+3)\, 4(2x^2+3)^3$$

$g(x) = 2x^2 + 3x$

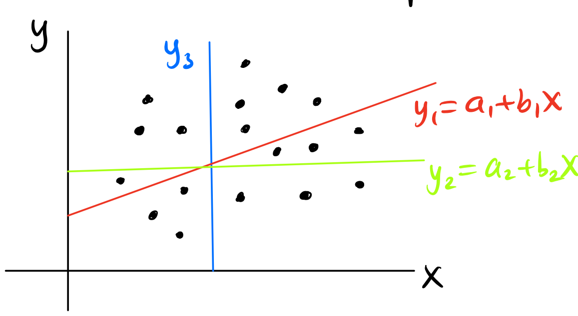
$g'(x) = 4x + 3$

$f(x) = x^4$

$f'(x) = 4x^3$

$f'(\underbrace{2x^2+3x}) = 4(2x^2+3x)^3$

Simple Linear Regression



$y_1 \rightarrow (a_1, b_1)$

$y_2 \rightarrow (a_2, b_2)$

$y_3 \rightarrow (a_3, b_3)$

Best line $\rightarrow$ Best (a, b)

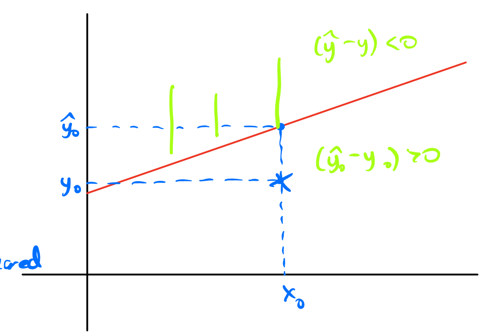
Quantify that y_1 is the best?

Loss Functions: how bad your model is:

$$\mathcal{L}(y_1) < \mathcal{L}(y_2) < \mathcal{L}(y_3)$$

$$\mathcal{L}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$= \text{MSE}$



sum of squared errors

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

pick parameters that give best model $\rightarrow$ lowest MSE

$$\hat{y}_i = a + bx_i$$

$\hat{y}$ = predicted value

→ linear reg

→ Neural network

→ Random Forest

$$\hat a, \hat b = \underset{a,b}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} (y_i - a - bx_i)^2$$

1. Pick a model
2. Pick loss
3. Find model minimises loss

Least Squares Derivation

- $\mathcal{L}(a,b) = \frac{1}{n}\sum_{i=1}^{n} (y_i - a - bx_i)^2$

$D = \{(x_i, y_i) ; i = 1, \dots, m\}$

- $\frac{\partial \mathcal{L}}{\partial a} = \frac{\partial}{\partial a}\frac{1}{n}\sum_{i=1}^{n} (y_i - a - bx_i)^2$

$= \frac{1}{n}\sum \frac{\partial}{\partial a}(y_i - a - bx_i)^2$

$= \cancel{n \times \frac{1}{n}} \sum_i (-1)(2)(y_i - a - bx_i) = 0 \times n$

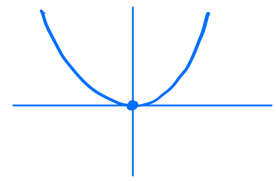
$\Rightarrow \sum_i (y_i - a - bx_i) = 0$

$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\underline{\sum_{i=1}^{n}} a - \frac{b}{n}\sum_{i=1}^{n} x_i = 0$

$\Rightarrow \bar y - a - b\bar x = 0$

$\Rightarrow \boxed{a = \bar y - b\bar x}$

$\hat X = \underset{X}{\operatorname{argmin}}\, X^2$



$f(x) = x^2$

$f'(x) = 2x = 0$

$\boxed{\hat x = 0}$

$\hat x = \underset{x}{\operatorname{argmin}}\, (2x+3)^2$

$f(x) = (2x+3)^2$

$f'(x) = 2(2)(2x+3) = 0$

$2x + 3 = 0$

$\hat x = -\frac{3}{2}$

$\sum_{i=1}^{n} 1 = n$

* $\bar f = \frac{1}{n}\sum_{i=1}^{n} f_i$

* $\overline{fg} = \frac{1}{n}\sum_{i=1}^{n} f_i g_i$

$\frac{\partial \mathcal{L}}{\partial b} = \frac{1}{n}\sum_i \frac{\partial}{\partial b}(y_i - a - bx_i)^2 = \frac{1}{n}\sum (-x_i)2(y_i - a - bx_i) = 0$

$\Rightarrow \frac{1}{n}\sum x_i y_i - a\frac{1}{n}\sum x_i - b\frac{1}{n}\sum x_i^2 = 0$

$\Rightarrow \overline{xy} - a\bar x - b\overline{x^2} = 0$

$\Rightarrow \boxed{b = \frac{\overline{xy} - a\bar x}{\overline{x^2}} \quad a = \bar y - b\bar x}$ Normal Equations

$\boxed{\overline{x^2}} \neq \bar x^2$

$b \neq f(a) = f(x,y)$

$b = \frac{\overline{xy} - (\bar y - b\bar x)\bar x}{\overline{x^2}}$

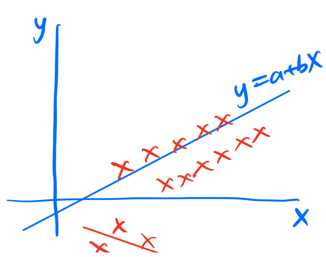
$= \frac{\overline{xy} - \bar x\bar y + b\bar x^2}{\overline{x^2}}$

$\Rightarrow b\left(1 - \frac{\bar x^2}{\overline{x^2}}\right) = \frac{\overline{xy} - \bar x\bar y}{\overline{x^2}} = \boxed{\frac{\overline{xy} - \bar x\bar y}{\overline{x^2} - \bar x^2}}$

$= \frac{\left(\frac{n}{n-1}\right)(\overline{xy} - \bar x\bar y)}{\left(\frac{n}{n-1}\right)(\overline{x^2} - \bar x^2)}$

$= \frac{\text{sample } \operatorname{cov}(x,y)}{\text{sample } \operatorname{var}(x)}$

$n\bar f = \sum f_i$



$b =$ slope of regression

$= \frac{\overline{xy} - \bar x\bar y}{\overline{x^2} - \bar x^2}$

Sample covariance:

$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar x)(Y_i - \bar y)$

$= \frac{1}{n-1}\sum_{i=1}^{n}(x_i y_i - x_i\bar y - y_i\bar x + \bar x\bar y)$

$$\text{Sample Variance} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$= \frac{n}{n-1}\left[\overline{x^2} - \bar{x}^2\right]$$

$$= \frac{1}{n-1}\left[n\overline{xy} - \bar{y}\sum x_i - \bar{x}\sum y_i + \bar{x}\bar{y}\sum_i^n 1\right]$$

$$= \frac{1}{n-1}\left[n\overline{xy} - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}\right]$$

$$= \frac{n}{n-1}\left[\overline{xy} - \bar{x}\bar{y}\right]$$

Multivariate Regression

y = House Prices

x_1 = #bdrms

x_2 = # bathrooms

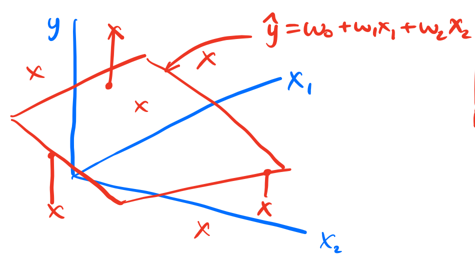
x_3 = suburb

x_4 = square feet

$\vdots$

x_p

- target / dependent
- predictors
- features
- covariates
- independent variables



$\hat{y}$ = predicted value

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \underbrace{(w_0 + w_1x_1 + w_2x_2 + \dots + w_px_p)}_{})^2$$

$$\hat{w}_0, \hat{w}_p, \dots \quad \underset{w_0, w_1, w_2, w_3, \dots w_p}{\operatorname{argmin}} \quad \frac{1}{n}\sum_{i=1}^{n}(y_i - w_0 - w_1x_1 - \dots - w_px_p)^2$$

$$\Rightarrow \boxed{\frac{\partial \mathcal{L}}{\partial w_0} = 0, \quad \frac{\partial \mathcal{L}}{\partial w_1} = 0, \quad \dots, \quad \frac{\partial \mathcal{L}}{\partial w_p} = 0}$$

$p+1$ normal equations

$$w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix} \in \mathbb{R}^{p+1} \qquad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$$

x_i = ith observation

$$= \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} \in \mathbb{R}^{p+1}$$

$$X = \begin{bmatrix} — & x^T & — \\ — & x_2^T & — \\ & \vdots & \\ — & x_n^T & — \end{bmatrix} \in \mathbb{R}^{n\times(p+1)}$$

$$= \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & & x_{2p} \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & x_{n1} & \cdots & & & x_{np} \end{bmatrix}$$

e.g. $v = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $z = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

$v^T = [1, 2]$

$v^Tz = 1\times 3 + 2\times 4$

$$\hat{y}_i = w_0 + w_1x_{i1} + \dots + w_px_{ip} = w^Tx_i$$

derivative when x is a matrix:

$$Cx \longrightarrow C^T$$

$$\underbrace{w^TCw}_{} \rightarrow 2Cw$$

$$\mathcal{L}(w_0, w_1, \dots, w_p) = \mathcal{L}(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^Tx_i)^2$$

$$Xw = \begin{bmatrix} x_1^Tw \\ x_2^Tw \\ \vdots \\ x_n^Tw \end{bmatrix}$$

$$= \frac{1}{n}\|y - Xw\|^2 \quad \text{(2-norm)} \quad \|\cdot\|$$

$$\|z\|^2 = \sum_{i=1}^{n} z_i^2 \qquad (Xw)^T = w^T X^T \qquad cw^2 \longrightarrow 2cw$$

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

$$\hat{w} = \underset{w}{\operatorname{argmin}} \frac{1}{n}(y - Xw)^T(y - Xw)$$

$$\mathcal{L}(w) = \frac{1}{n}(y - Xw)^T(y - Xw) = \frac{1}{n}\left[y^Ty - y^TXw - w^TX^Ty + w^TX^TXw\right]$$

$$= \frac{1}{n}(y^Ty - \underbrace{2y^T}Xw + w^T\underbrace{X^T}Xw)$$

$y^TXw - w^TXy$ $\qquad y^T \in \mathbb{R}^{1\times n}$

$= y^TXw - (y^TXw)^T$ $\qquad X \in \mathbb{R}^{n\times(p+1)}$

$= 2y^TXw$ $\qquad w \in \mathbb{R}^{(p+1)\times 1}$

$y^TXw \in 1\times 1$

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{1}{n}(0 - 2X^Ty + 2X^TXw) = 0$$

$$\Rightarrow X^TXw - X^Ty \Rightarrow w^T = (X^TX)^{-1}X^Ty$$

$p = 1$ (simple linear regression) $\quad w = (X^TX)^{-1}X^Ty$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_p \end{bmatrix} \leftarrow \text{1st house} \quad \in \mathbb{R}^{n\times 2} \qquad y \in \mathbb{R}^n \qquad w = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} \in \mathbb{R}^2$$

- $(X^TX) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_p \end{pmatrix}\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_p \end{pmatrix} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & n\bar{x^2} \end{pmatrix}_{2\times 2}$ $\quad n\bar{x}$: $\sum x_i$

$\in \mathbb{R}^{2\times n}$ $\mathbb{R}^{n\times 2}$

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad A^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

- $(X^TX)^{-1} = \dfrac{1}{n^2(\bar{x^2} - \bar{x}^2)}\begin{pmatrix} n\bar{x^2} & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix}$

- $X^Ty = \begin{pmatrix} n\bar{y} \\ n\bar{xy} \end{pmatrix}$

- $w = (X^TX)^{-1}X^Ty = \begin{pmatrix} w_0 \\ \hat{w}_1 \end{pmatrix} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$