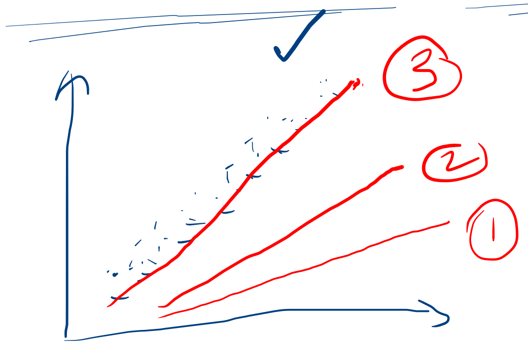
Linear regression using gradient descent:



$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \cdots \theta_n x_n$

Step 1

$h_\theta(x) = \theta_0 + \theta_1 x$

at start

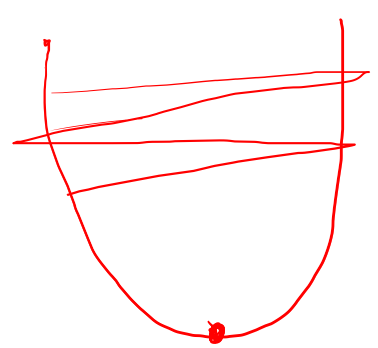
$\theta_0 = 1$ $\theta_1 = 1$ (random initialization)

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x) - y)^2$$

Step 2

compute error caused by this line

$\frac{\partial}{\partial \theta} J(\theta) \rightarrow$ error term



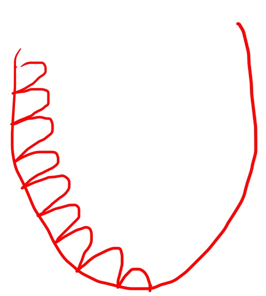
Iteration

Step 3

update the θ value

$\alpha \rightarrow$ learning

$$\theta_{new} = \theta_{old} - \alpha \frac{\partial}{\partial \theta} J(\theta)$$



Step 4

repeat step 3 & step 4

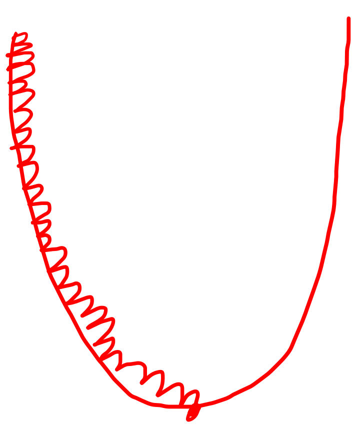
Saturation

↳ stop the algorithm.

iteration.

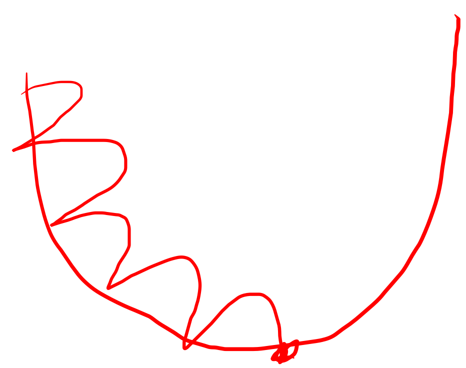
α -> Learning

α -> Low Learning



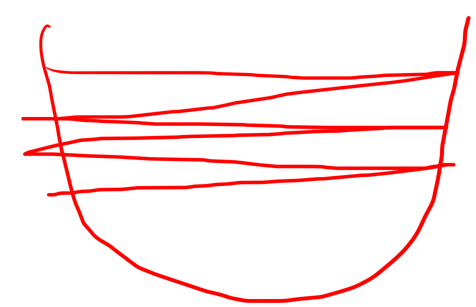
↳ local minima (or) global minima

α -> optimum



↓ Optimum steps to reach minima

α -> large



Moves & misses the minima

gradient descent

slope

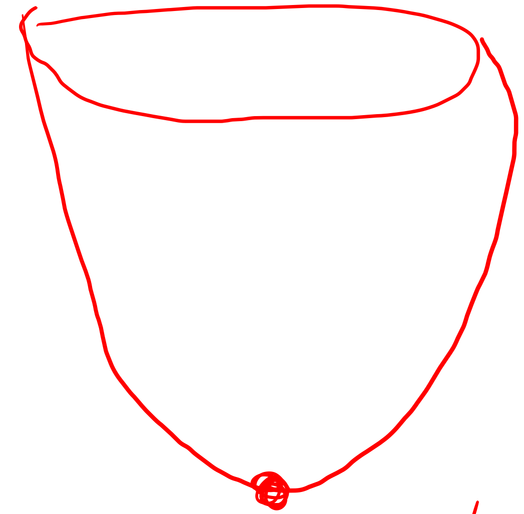
moving downward.

local minima

global minima

Linear regression

convex curve. / convex optimization



global / local minima are same.

gradient descent

$$\left\{ \theta_{new} := \theta_{old} - \alpha \frac{\partial J(\theta)}{\partial \theta} \right\}$$

$h(\theta) = \theta_0 + \theta_1(x_1) \ldots$

$$\left\{ \begin{array}{l} \rightarrow \theta_0 := \theta_{0\,old} - \alpha \dfrac{\partial J(\theta)}{\partial \theta_0} \\ \rightarrow \theta_1 := \theta_{1\,old} - \alpha \dfrac{\partial J(\theta)}{\partial \theta_1} \\ \vdots \end{array} \right\}$$

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{\partial}{\partial \theta_0}\left[\frac{1}{2m}\sum_{i=1}^{m}\left(\underline{(\theta_0+\theta_1 x)}-y\right)^2\right]$$

$$\frac{d}{dx}x = 1$$

$$= \frac{\cancel{2}}{\cancel{2}m}\sum_{i=1}^{m}\left((\theta_0+\theta_1(x))-y\right)\cdot(1)$$

$$\boxed{\frac{\partial}{\partial \theta_0} = \frac{1}{m}\sum_{i}^{m}\left(h_\theta(x)-y\right)}$$

$$\frac{\partial J(\theta)}{\partial \theta_1} = \frac{\partial}{\partial \theta}\frac{1}{2m}\sum_{i=1}^{m}\left(\underline{(\theta_0+\theta_1 x_1)}-y\right)^2$$

$$\frac{d}{d\theta_1} = \cancel{\theta_0}+\theta_1 x$$

$$\frac{d}{d} = x$$

$$\boxed{\frac{\partial J(\theta)}{\partial \theta_1} = \frac{2}{2m}\sum_{i=1}^{m}\left(h_\theta(x)-y\right)x}$$

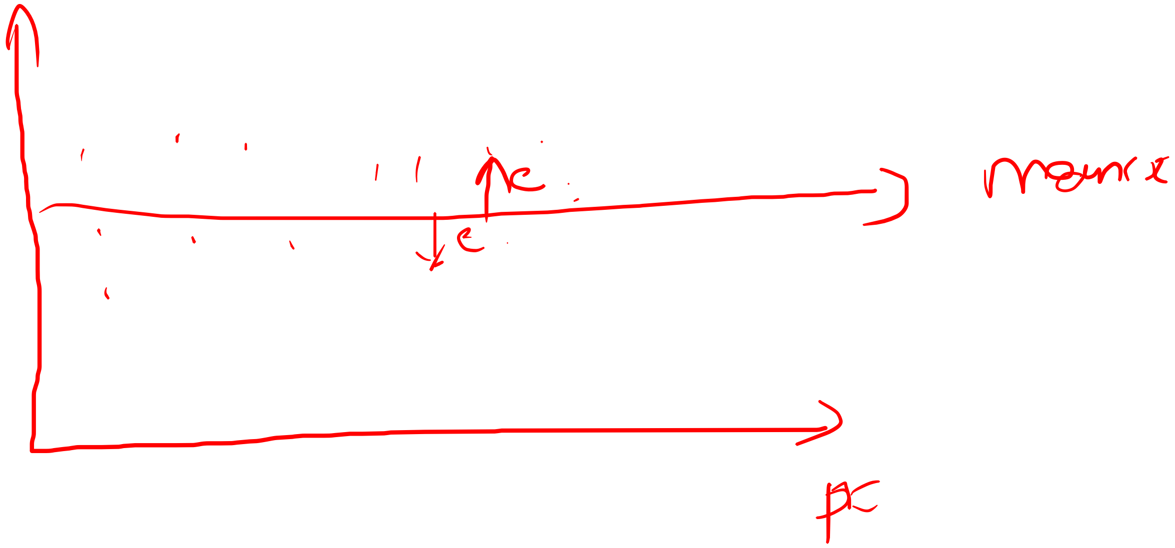
gd for $\theta_0 + \theta_1 x$

$$\left\{ \begin{aligned} \theta_0 &= \theta_{0_{old}} - \alpha \frac{1}{m} \sum (h_\theta(x) - y) \\ \theta_1 &= \theta_{1_{old}} - \alpha \frac{1}{m} \sum (h_\theta(x) - y)\, x_1 \end{aligned} \right\}$$

Ex

$$\frac{d}{dx}(x+1)^2 = 2(x+1)^{\times 1}$$

$$= 2(x+1)$$



Measuring The errors

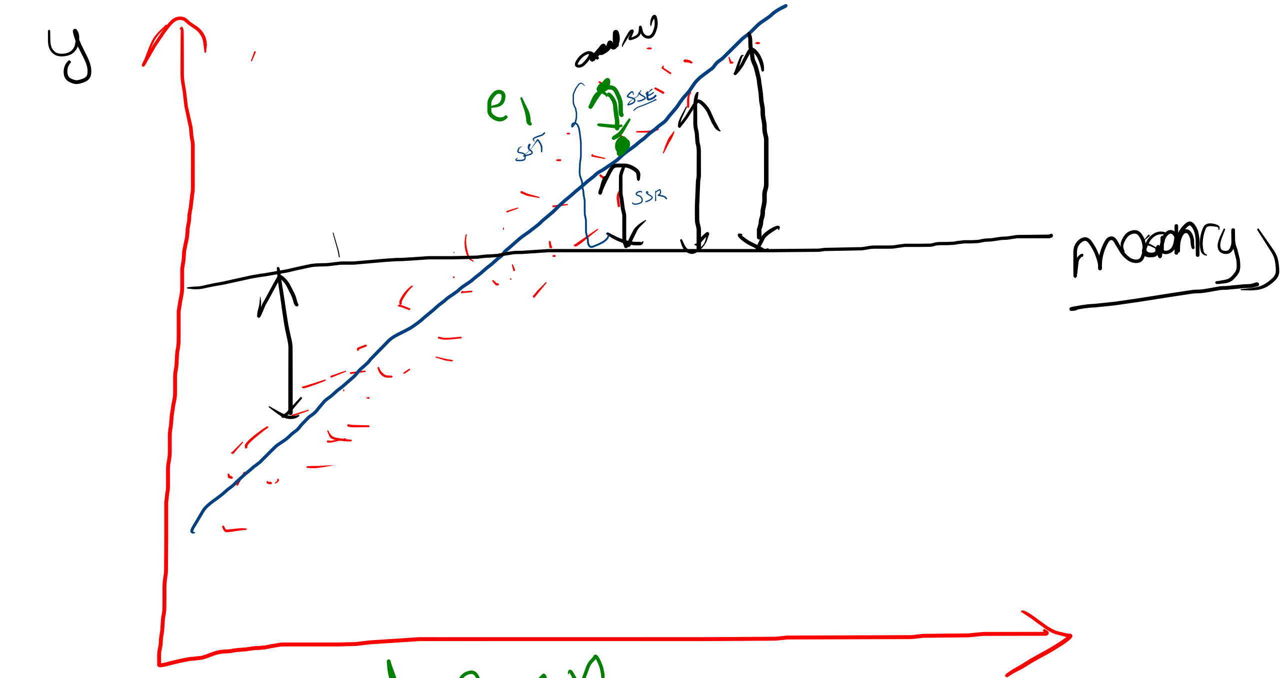
(or)

how accurate is my algorithm

SSE = SST

$$\boxed{SSE = \sum (\text{predicted} - \text{actual})^2}$$

$$= \sum_{i=1}^{m} e_i^2$$



Sum of squared error

$$\boxed{SSE = \sum (y - \hat{y})^2} \rightarrow (1)$$

$$\boxed{SST = SSE + SSR}$$

SSR
↳ sum of regression

$$\boxed{SSR = \sum (\bar{y} - \hat{y})^2} \rightarrow (2)$$

$$\boxed{SST = \sum (y - \bar{y})^2} \rightarrow (3)$$

(1) R^2

Coefficient of determination $= \frac{SSR}{SST}$ (or) $1 - \frac{SSE}{SST}$

0 — 1

↓ poor linear algo

↓ accurate algorithm

(2) mean squared error

$SSE = \Sigma(y-\hat{y})^2$

$$\boxed{MSE = \frac{SSE}{n}}$$

(3) RMSE

$RMSE = \sqrt{MSE}$

(4) $MAE = \frac{\Sigma|y-\hat{y}|}{n}$

$\frac{6-4}{6} = \frac{2}{6} = 0$

$(6-4)$ 2

(5) MAPE =

$$MAPE = \left(\frac{1}{n}\Sigma\left|\frac{y_1-\hat{y}}{y_1}\right|\right)\times 100$$

(6) MPE

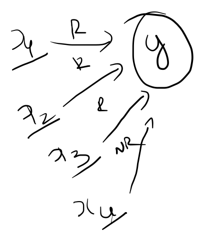
$$= \left(\frac{1}{n}\Sigma\frac{(y_1-\hat{y})}{y_1}\right)\times 100$$

(7) Adj R^2 ||

R^2

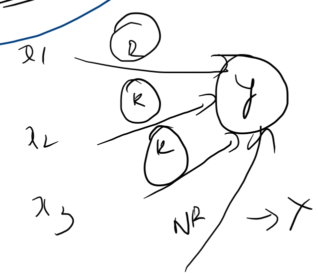
$R^2 = \frac{SSR}{SST}$ (or) $1 - \frac{SSE}{SST}$

$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \ldots$ $\uparrow$



$R^2 \uparrow$ (not won't consider whether are related to my y)

Adj - R^2



x_4

Adj - R^2 = $\uparrow$ only when you add related variable

won't increase when you add variable which are not related

Add penalty to R2

$$\text{Adj} = R^2 - \frac{k-1}{n-k}(1-R^2)$$

penalty 99%

$h_\theta(x) = \theta^T x$

$R2 \to 0.76$

$K \to 5$

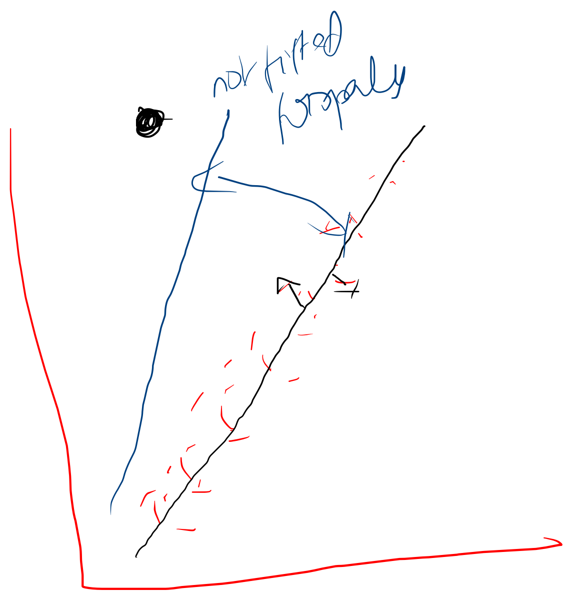
$n \to 1000$

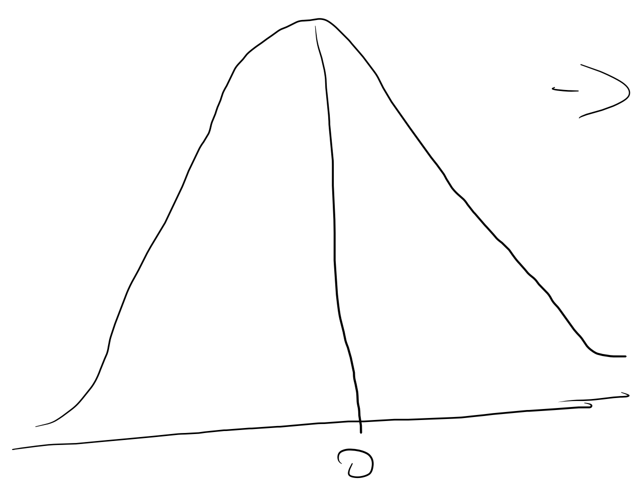
$$= 0.76 - \left(\frac{4}{995}\right)(0.24)$$

=

Assumption of linear regression

① linear regression → $y \sim x$

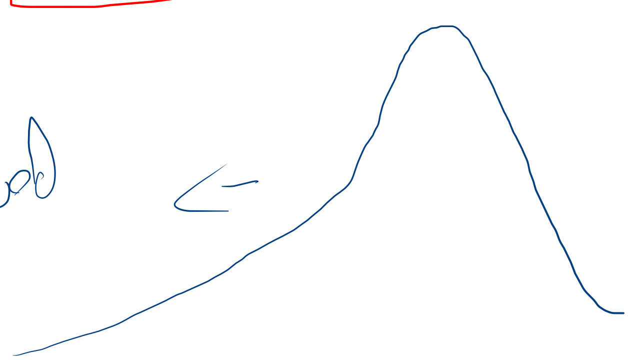
② 



→ normality of error term

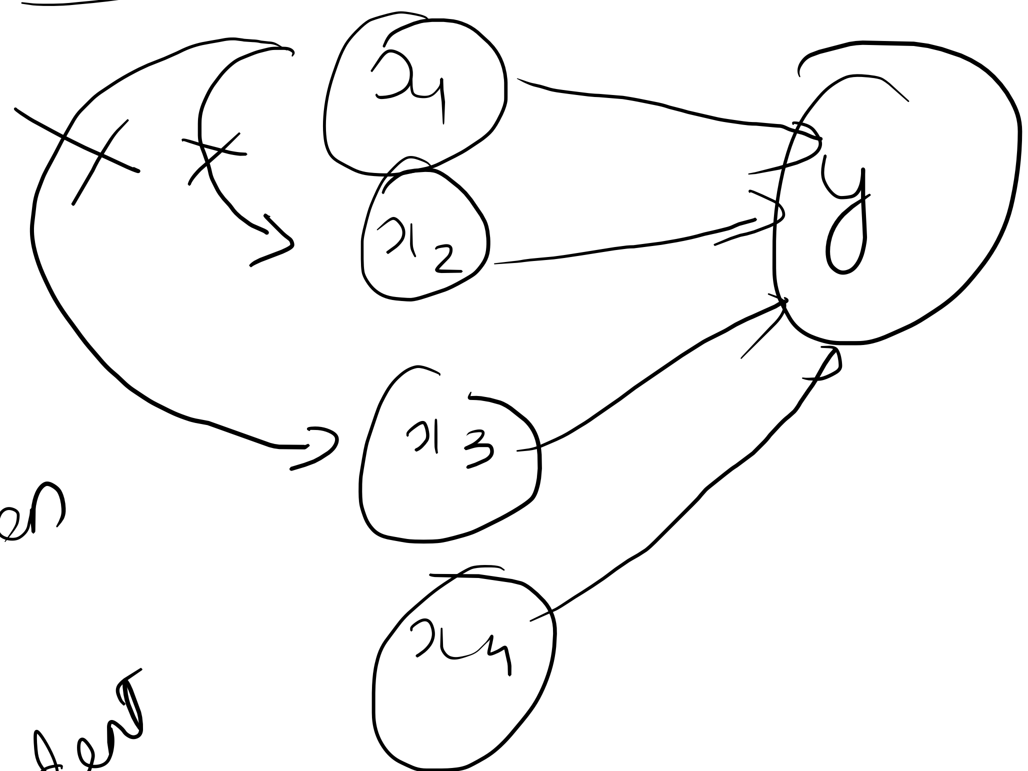
(or)

residue should be normally distributed

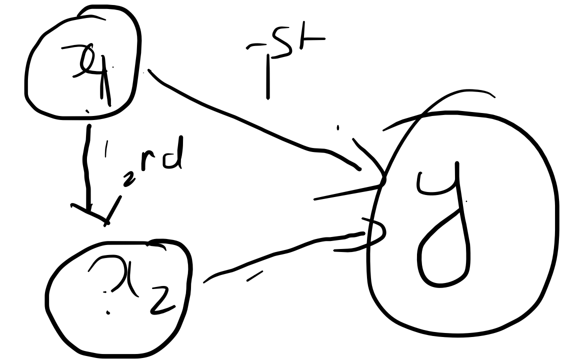


Skewed when outlier

3) multicollinearity



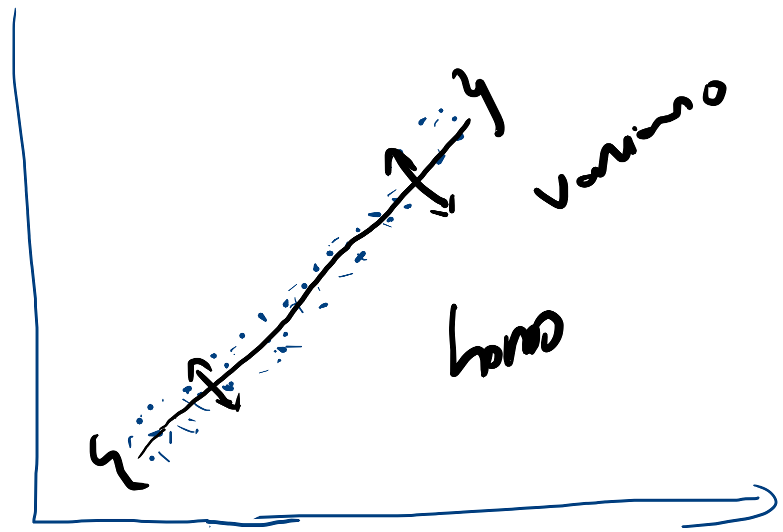
no relation
x
(independent

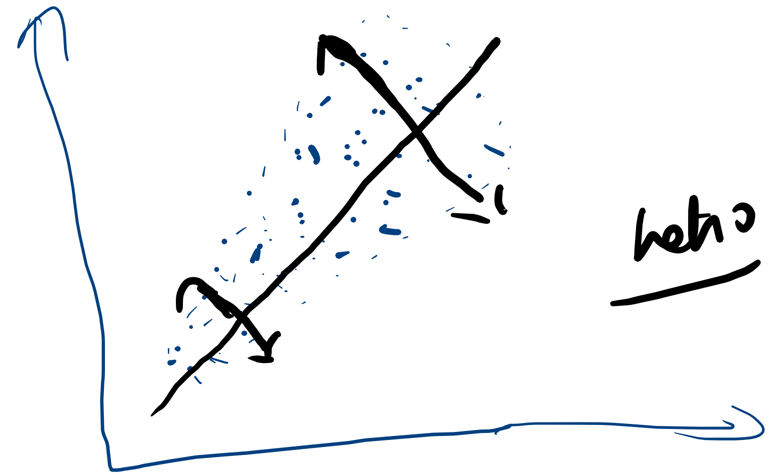


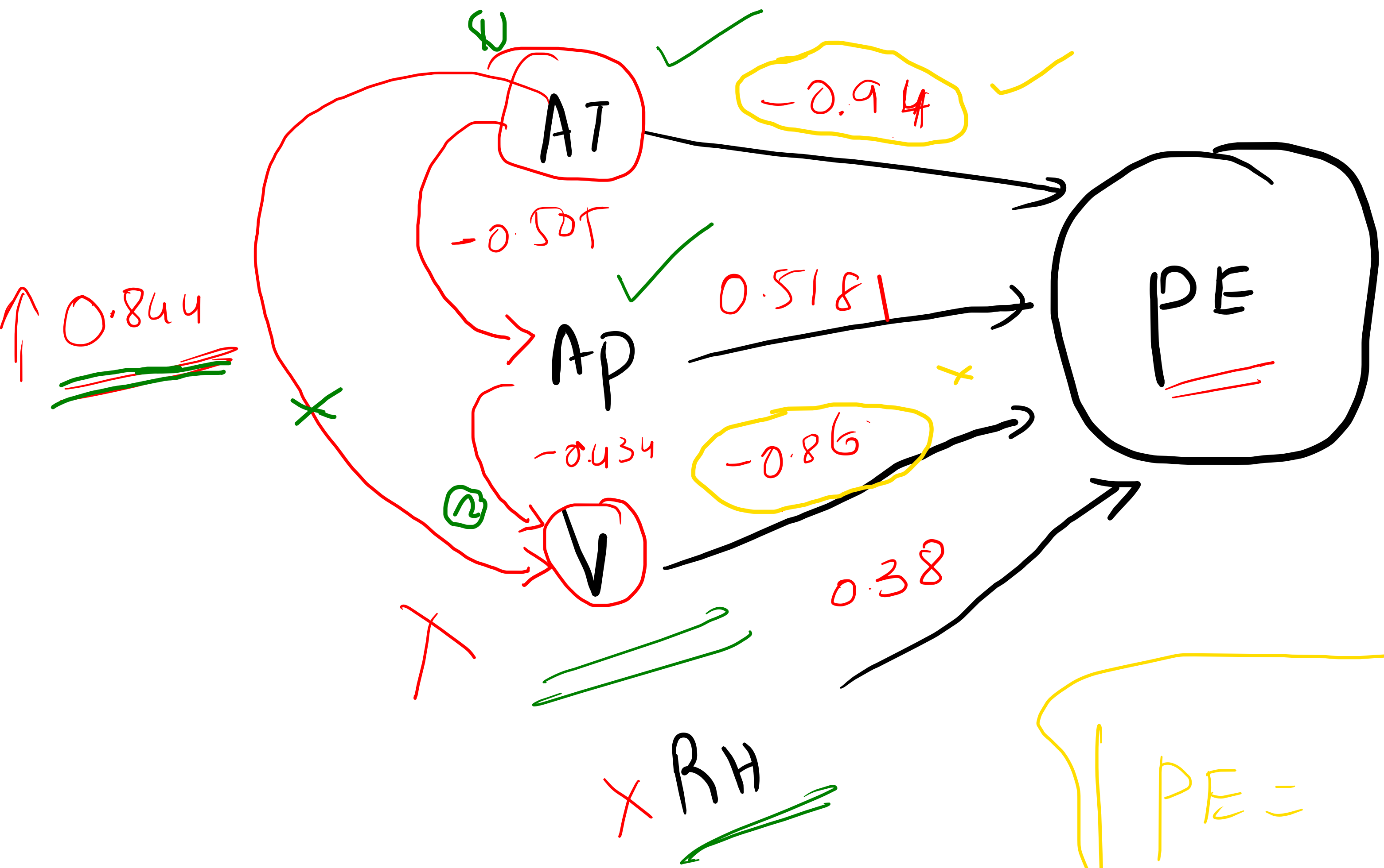
multicollinearity → corr

corr

variance inflation factor







± 0.5

$$PE = 349.57 - 2.11\,AT + 0.144\,Ap$$

Normal equation

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