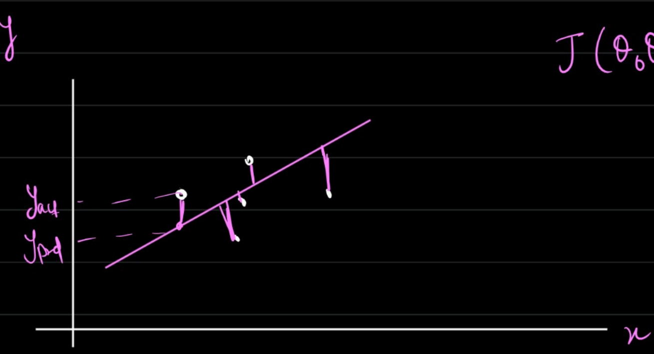
Evaluation metrics for Regression problem (MSE, MAE, RMSE)

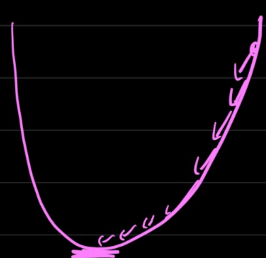
① MSE (Mean squared Error)



$$J(\theta_0 \theta_1) = \frac{1}{n} \underbrace{\sum_{i=1}^{n} (y_i - h_\theta(x))^2}$$

$$\text{Error} = \frac{1}{n}\sum_{i=1}^{n} \underbrace{(y_{act} - y_{pred})^2}_{\text{squared Error}}$$

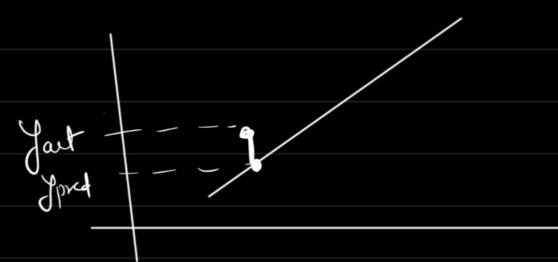
Mean



* Idea → how close our predictions are to the actual value.

→ Error is quantified

→ Lower the MSE better the model will be.



* MSE measures the average of the squared differences between the predicted and actual values.

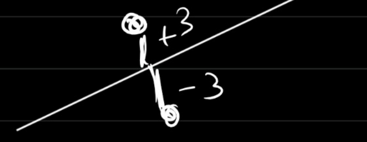
$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_{act} - y_{pred})^2$$

↳ quantitative measure how well the model's prediction align with True out comes.

Why square ?

Why not directly $y_{act} - y_{pred}$

* Squaring the differences ensures that positive and negative error don't cancel out each other.



$+3-3=0$

$(+3)^2+(-3)^2$

* Advantages of MSE

→ Errors don't cancel out each other.

→ MSE used as CF.

→ It is differentiable.

(2) $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{act}-y_{pred})^2$

→ It is a convex fn.

It has only one local and one global minima.

Gradient descent:

$m_{new} = m_{old} - \eta \frac{\partial CF}{\partial m_{old}}$

→ Emphasis on large Errors ⇒ squaring amplifies impact of large error, sensitive to significant deviation

$\frac{\partial x^n}{\partial x} = n x^{n-1}$

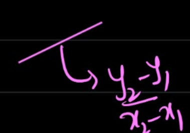
→ Quadratic eqn

Convex function

↓

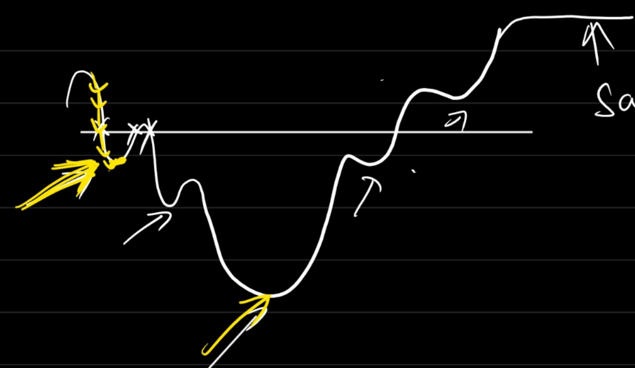
One minima

↓



CF here as only one minima, simple linear Regression.

Non Convex Cost function

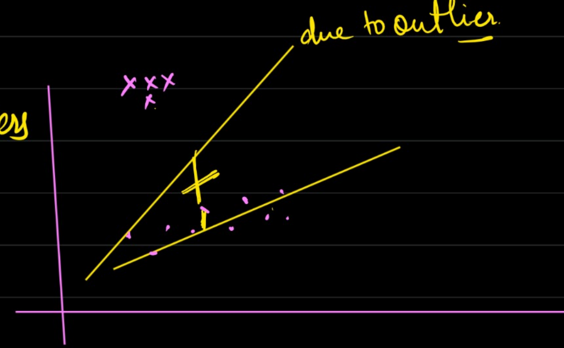


* Disadvantage of MSE

→ Not robust to Outlier.

$(y_{act} - y_{pred})^2 \uparrow\uparrow$

due to outliers



→ It is not in the same unit

Ht	wt (kg)
	–
	–
	–
	–
	–

$(y_{act} - y)^2$

$y_{act} \to$ kg $\to y_{pred}$

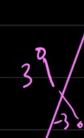
$15\,kg - 10\,kg = (5\,kg)^2$

$\Rightarrow 25\,(kg)^2$

② Mean Absolute Error

→ It measures the absolute difference b/w the actual and predicted value.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_{act} - y_{pred}|$$



$|3| + |-3|$

$3 + 3 = 6$

Advantage

→ MAE does not square the differences, making it less sensitive outliers. Each absolute difference contribute equally to outlier.

→ It will be of same unit

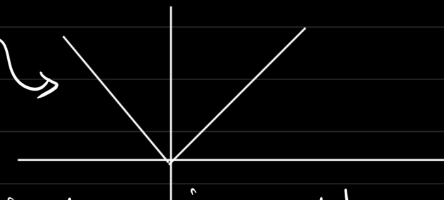
$|5m - 4m|$

$1m$

→ It is more interpretable.

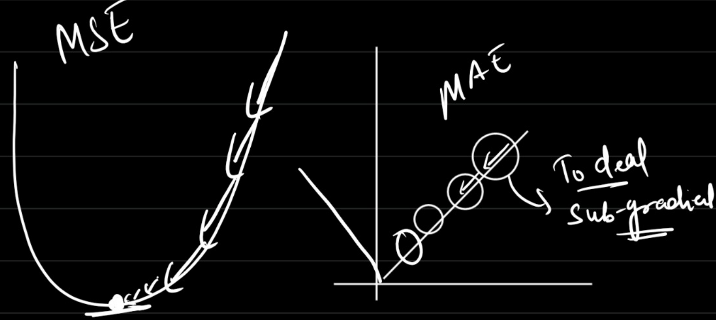
* Disadvantage

$|x|$



① Convergence usually takes time. optimization is complex

② Time consuming.



MSE = 

↓

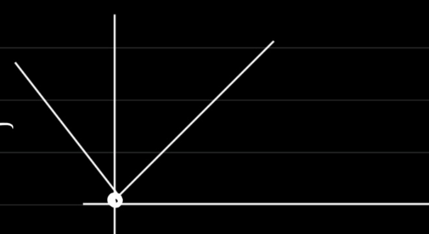
Twice differentiable

but

MAE ~~is once differentiable.~~

$$\boxed{y_{act} = y_{pred}}$$

$|x|$ ←



↓

differentiable everywhere except $x=0$

③ RMSE (Root Mean Squared error)

$$RMSE = \sqrt{MSE}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_{act} - y_{pred})^2}$$

$$\sqrt{(5kg - 3kg)^2}$$

$$\sqrt{(4kg)^2} = 2kg$$

Advantage

→ Same unit

→ Differentiable.

→ less sensitive to outliers

→ Magnifies to large error.

* disadvantage

→ interpretation becomes complex

✓ R square & adjusted r square :- Evaluation metrics

✓ MSE , MAE, RMSE → Evaluation metric / cost function.