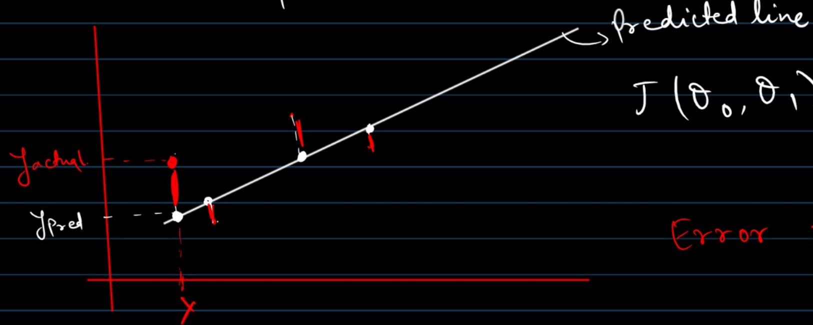
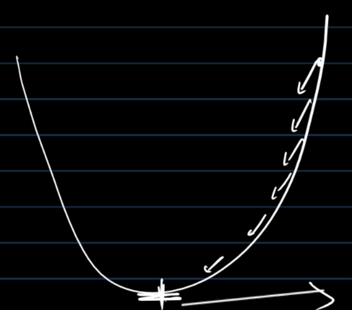
* Evaluation Metrics for Regression (MSE, MAE, RMSE)

① MSE (Mean squared Error)



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

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



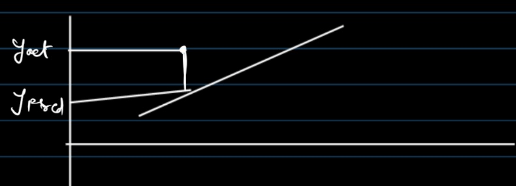
idea :- how close our predictions are to the actual value.

Why ? → For best fit line

To come to the minima value, we need to minimize the error.

* How ? → by Quantifying the Error.

* Lower the MSE, better the model will be.



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

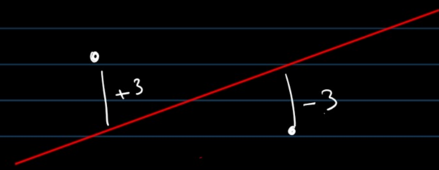
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(y_{act} - y_{pred} \right)^2$$

↳ quantitative measure how well the model prediction align with the true outcomes.

* 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.

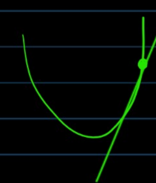
→ Errors don't cancel out each other.

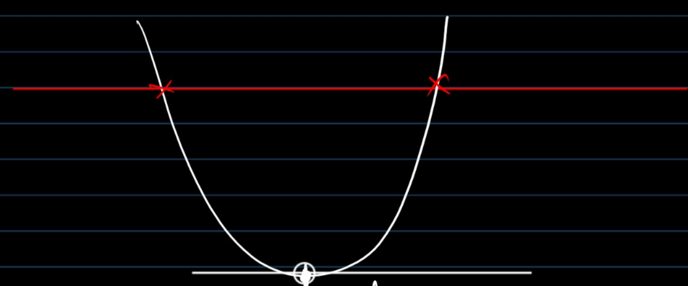
→ MSE is used as CF → $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{act} - y_{pred})^2$

→ MSE is differentiable → $\theta_j := \theta_j - \eta \frac{\partial CF}{\partial \theta_j}$

→ It is a convex function. It has only one local and one global minima.

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





Convex function → One minima

Cost fn has only one minima.

* non-convex cost function.

→ Emphasis on large Error: Squaring amplifies the impact of large error ⇒ Sensitive to significant deviation.

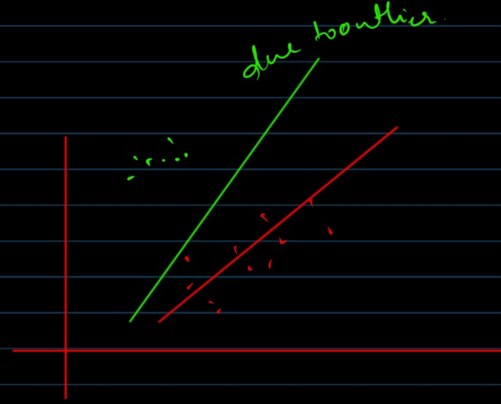
$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{act} - y_{pred})^2$

$= (5-3)^2 \qquad (10-3)^2$

$= 4 \qquad 7^2 = 49$

* Disadvantage

→ Not robust to Outlier.



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

→ It is not in the same unit

Height (X)	(Y) weight (kg)

$$(y_{act} - y_{pred})^2$$
$$(\underline{10kg} - \underline{5kg})^2$$
$$(5kg)^2$$
$$\underline{25kg^2}$$

② Mean Absolute Error

→ It measures the absolute difference between the actual & Predicted value.

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

$$|3| + |-3| = 3+3 = 6$$

* Advantage

→ MAE doesnot square the differences, making it less sensitive to outliers. Each absolute value difference contribute equally to outlier.

→ It will be of same unit $|5m - 4m| = 1m.$

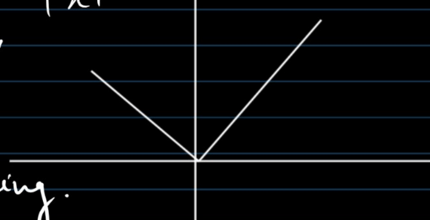
→ It is more interpretable

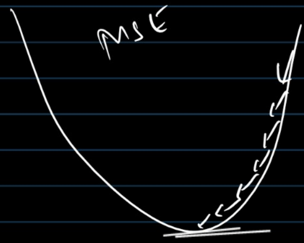
* Disadvantage

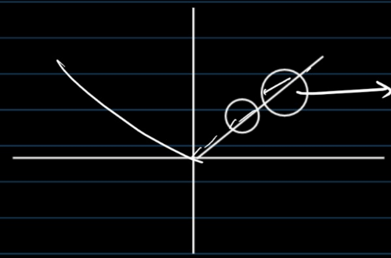
① Convergence usually takes time. Optimization is complex

② Time consuming model training.

$|x|$



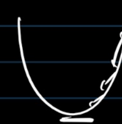




To deal

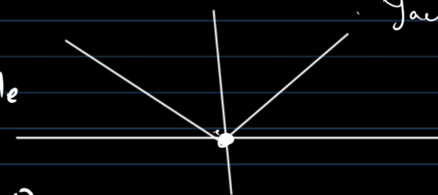
Sub-gradient

MSE =



Twice differentiable

MAE



$|x|$ is differentiable except $x = 0$

③ RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{MSE}$$

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

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

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

Advantage

→ Same unit

→ Differentiable

→ less sensitive to Outliers

→ Magnifies the large error

Disadvantage

→ Interpretation becomes complex.

Rsquare & adj rsquare — Evaluation metrics.

MSE, MAE, RMSE → Evaluation metrics / Cost function.