Difference b/w cost & Loss function

m = no. of data points.

Cost function like MSE = $J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)} \right)^2$

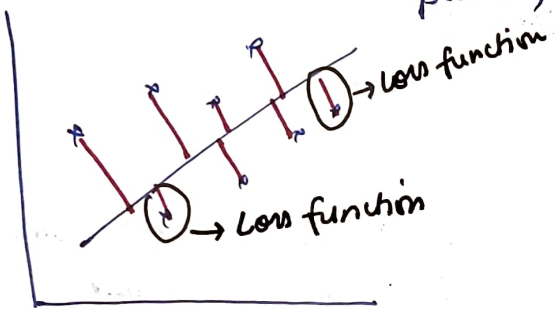
Cost function is for all data points together.

$h_\theta(x)^{(i)}$ ↓ predicted, $y^{(i)}$ ↓ Actual

Loss function = $\left(h_\theta(x)^{(i)} - y^{(i)} \right)^2$

in loss function, we will not do summation of all points. Loss function → for every single data point, find the difference.

- Av is not taken.
- Summation is not taken.



Cost function

① MSE ② MAE ③ RMSE ④ Huberloss

① MSE advantages -

Disadvantage

- Not robust to outliers.

- Our main aim is to get one global minima.
- MSE gives one global minima.
- MSE is a quadratic equation.
- With quadratic equation we always get convex function with 1 global minima & no local minima

* there are few equations which gives many local minima also where algorithm thinks we already reached minima & it gets stuck there but in actual its not global minima. In that situation we have to convert it into 1 global minima.

- Non-convex to convex function.