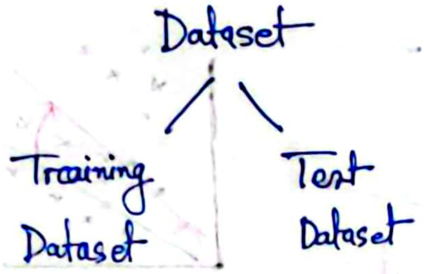
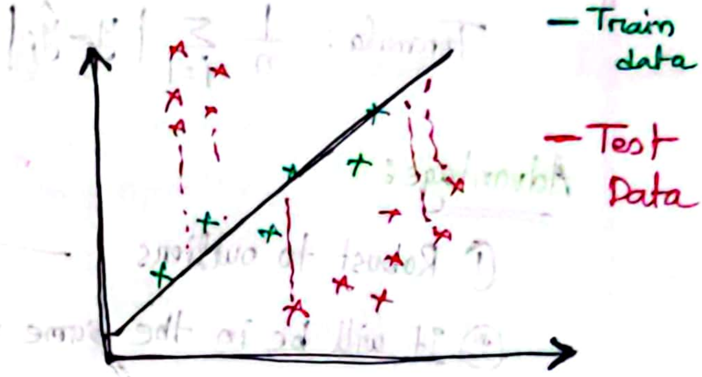
So, it should be better to use MAE when you have outliers and use ~~MA~~ MSE or RMSE when you don't have outliers.

Ridge Regression: (L2 Regularization) → $\boxed{\text{Reducing Overfitting}}$

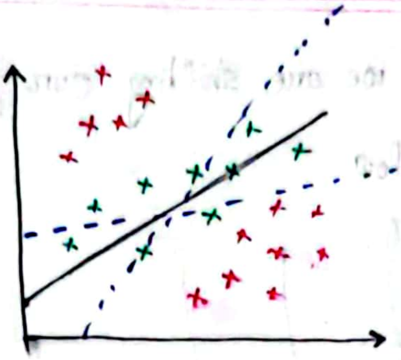




In the above plot, we could see that, for training dataset, model train very well because the error (cost function) is less but for test data we can see that, the errors the very high. This incident called 'Overfitting'.

$$\left.\begin{array}{l}\text{Train data} \rightarrow \uparrow\uparrow \text{ Accuracy} \\ \text{Test data} \rightarrow \downarrow\downarrow \text{ Accuracy}\end{array}\right\} \text{Overfitting}$$

The aim of Ridge Regression is to balance the error by modifying the fit line. It doesn't make the error 0, rather minimize it.

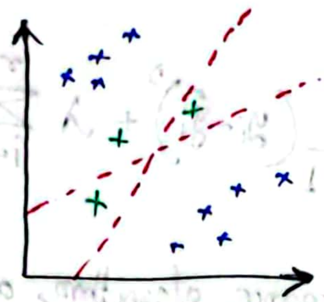


To minimize the error, we have to modify the cost function.

Modified Cost Function: $\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y})^2 + \boxed{\lambda \sum_{j=1}^{n}(\underset{\theta_1}{\text{slope}})^2}$

λ = Hyper parameters.

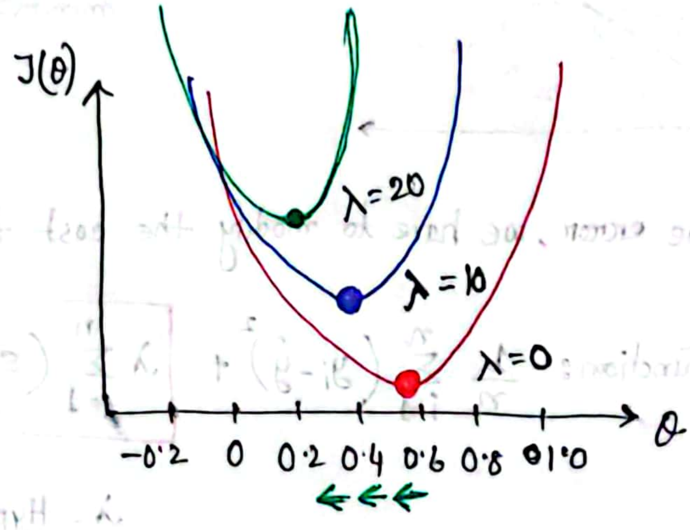
For, adding extra parameters, cost function (error) will never be 0 means the model will be never accurate. The best fit line will shift from it's accurate position.



For the case of multiple linear regression:

modified cost function: $\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y})^2 + \lambda \sum_{i=1}^{n}\left[(\underset{\theta_1}{\text{Slope}})^2 + (\underset{\theta_2}{\text{Slope}})^2 + (\underset{\theta_3}{\text{Slope}})^2\right]$

By adding λ (hyper parameter) we are shifting our gradient descent of the cost function like below.



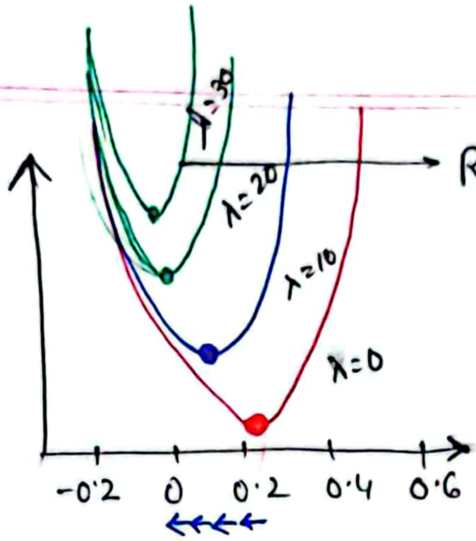
② Lasso Regression (L1 Regularization) → Feature Selection

cost function : $\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{n} |\text{slope}_{\theta_1}|$

for multiple linear regression:

cost function: $\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{n} \left[|\theta_1| + |\theta_2| + |\theta_3| \ldots + |\theta_n| \right]$

Now, During using the convergance algorithm, as much as λ will be increasing the global minima of the least correlated feature will turn into 0 value. So, at that time we can understand that feature is no longer required and we can delete it.



Remove this feature because this the least correlated with target feature.

$\lambda \uparrow\uparrow\uparrow$

$\theta \downarrow\downarrow\downarrow$ and beome 0

When $\theta = 0$, remove that feature

Elastic Net Regression:

→ Reducing overfitting → Ridge

→ Feature Selection → Lasso

In this regression, we are combining both concepts to the formula.

cost function:
$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}_{\text{MSE}} + \underbrace{\lambda_1\sum_{i=1}^{n}(\underset{\theta_1}{\text{slope}})^2}_{\text{Reduce Overfitting}} + \underbrace{\lambda_2\,|\underset{\theta_1}{\text{slope}}|}_{\text{Feature selection}}$$

$(\lambda_1, \lambda_2) \rightarrow$ { Hyperparameter tuning }