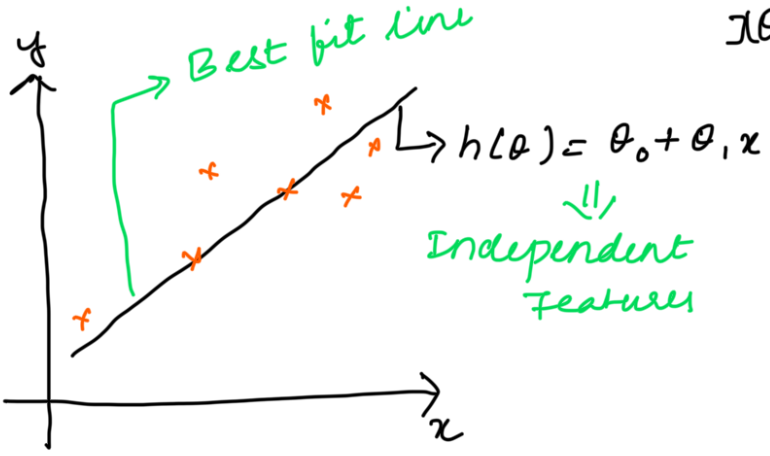
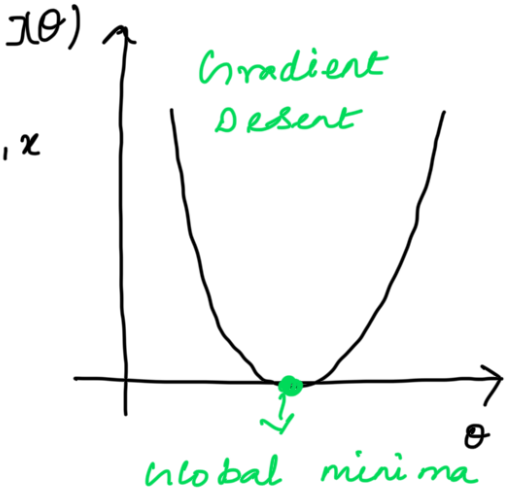
Ridge Regression, Lasso Regression, Elastic net Regression

Linear Regression





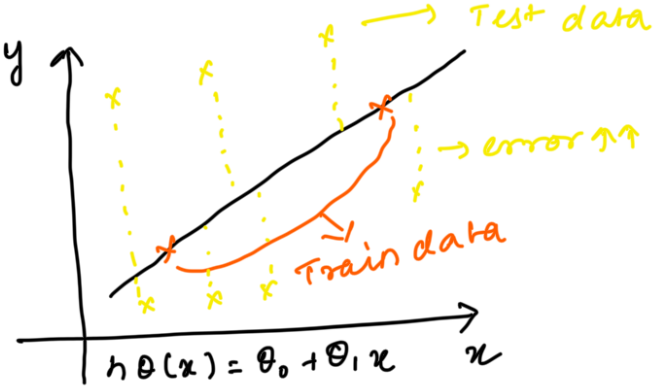
$$\text{cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^i - y^i\right)^2$$

$\Downarrow$

mean squared error

① Ridge regression

Let say I have the below data points



Over fitting

Train data → accuracy ↑ → low Bias

Test data → accuracy ↓ → low Variance.

here my model is over fitting. here the model performed well with the train data, however with the test data, error increases, this the problem of

Note:

we should not get 100% accuracy on

Overfitting.

the training data, as this will leads to overfitting.

Ridge regression is also called as L2 Regularization which helps to reduce overfitting

Suppose, my linear regression has some overfitting. in order to reduce the overfitting in linear regression we use Ridge regression.

We can consider the Ridge regression as a new algorithm to hyperparameter tune the Linear regression.

Cost fn of Ridge regression

$$\text{cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^i - y^i\right)^2 + \lambda \sum_{i=1}^{n} (\text{slope})^2$$

Hyper parameter.

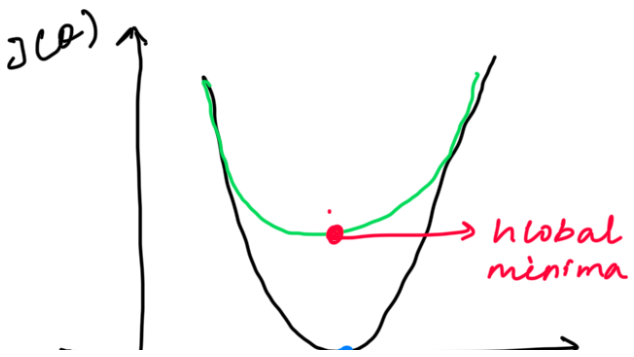
Let $\lambda = 1$,

$$= 0 + 1\left[(\theta_1)^2\right]$$

when adding these two parameters, the situation is going to be in such a way that, not going to get the best fit line that passes exactly on my traing data.

∴ my cost function will always > 0

This is how ridge regression helps to avoid overfitting.



$\lambda = 0$

$\lambda \neq 0$

to when ever, λ value changes gradient descent

will also change, will also get new global value, and θ value reduces.

how Ridge regression reducing overfitting?

→ It reducing the input by reducing the coefficient of the feature that are not directly releated to the output feature.

② Lasso Regression

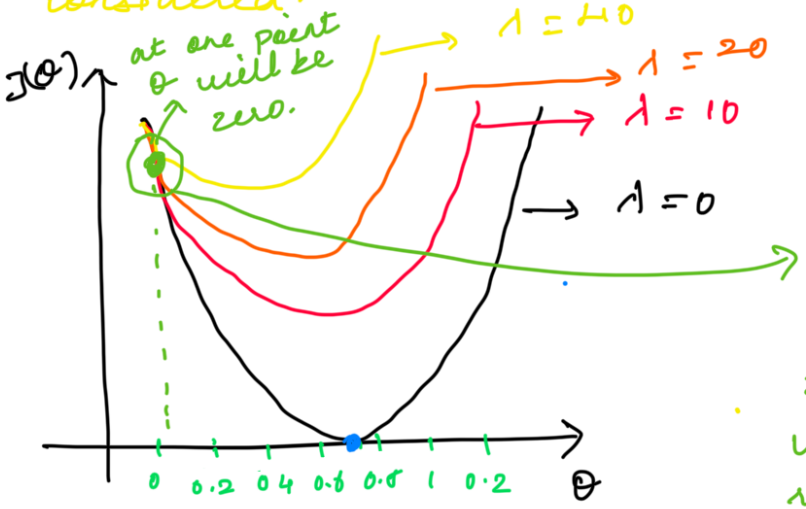
It is also called as L1 regularization, we specifically use Lasso regression for feature selection.

$$\text{cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^i - y^i\right)^2 + \lambda\sum_{i=1}^{n}|\text{slope}|$$

used for feature selection.

What is feature selection?

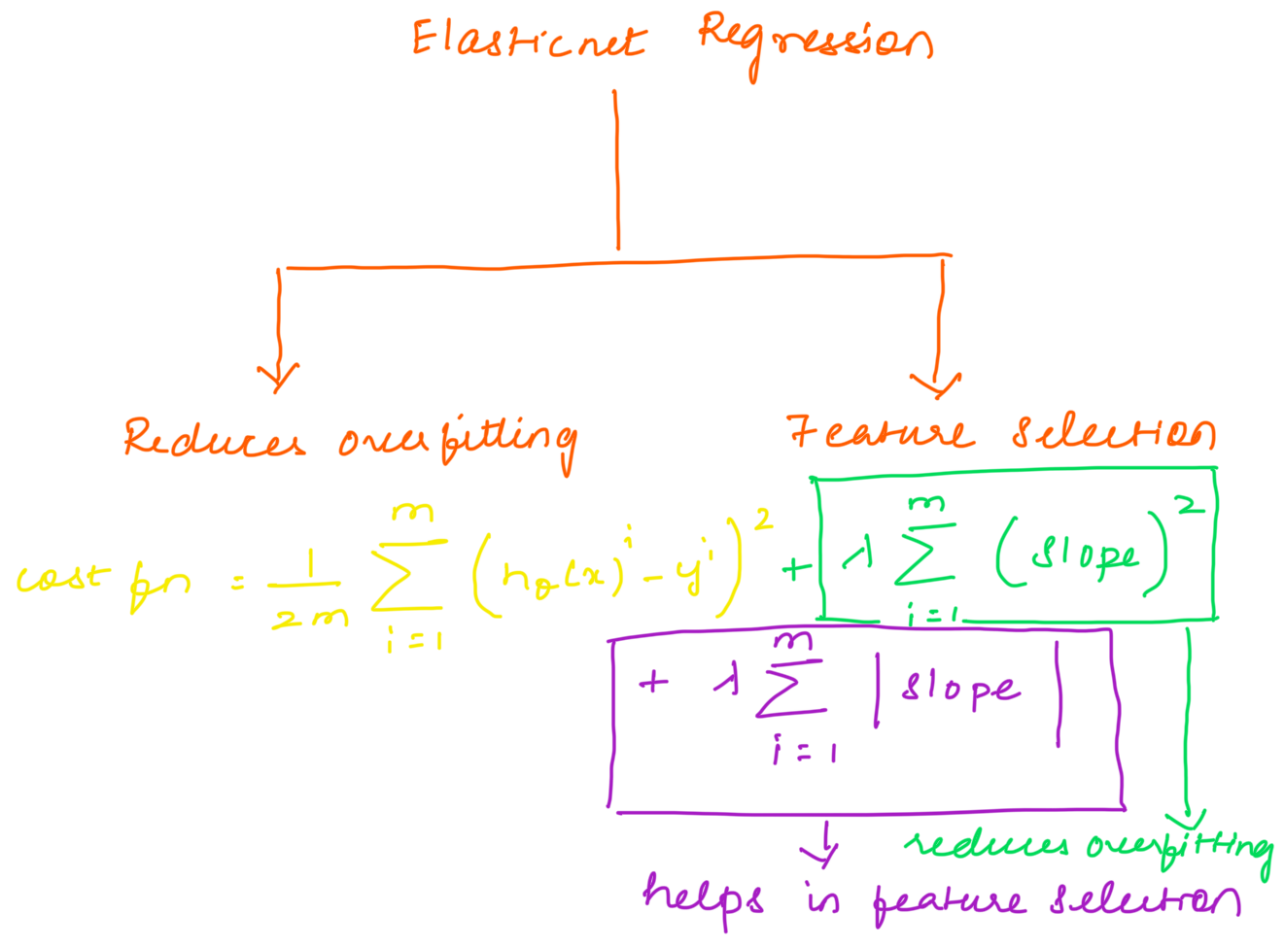
The feature that are not that important will be deleted and the features that are important will be considered.



θ value becoming zero → coefficient is actually becoming zero. → In short we are trying to remove the specific feature.

③ Elasticnet Regression

→ It is the combination of Lasso and Ridge regression



Why we are using Ridge, Lasso and Elasticnet regression

→ Hyperparameter tuning the Linear regression.