Regularization

This is a form of regression that constraint / regularizes or shrinks the coefficient estimates towards zero. This technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting. They have two techniques lasso regularization / L1 and ridge regularization / L2.

When it comes to training models, there are two major problems one can encounter:

Overfitting happens when the model performs well on the training set but not so well on unseen (test) data.

Underfitting happens when it neither performs well on the train set nor on the test set.

Lasso Regression

It is a type of linear regression that includes a penalty term, which is the sum of the absolute values of the coefficients. This penalty term encourages sparsity in the model, meaning it can shrink some coefficients to exactly zero. This makes Lasso useful for feature selection as well as for reducing model complexity and preventing overfitting.

In short, this is a regularization technique used in feature selection using a shrinkage method. In lasso regularization magnitude of coefficients can exactly zero.

$$\text{Cost function} = \text{loss} + \lambda \sum ||w||$$

Loss – sum of squared residual, λ – penalty, w – slope of curve

$$mean\ square = \frac{1}{n} \sum (y - \bar{y})^2 + \lambda\,|\mathrm{w}|$$

Ridge Regression

It is another type of linear regression that includes a penalty term, but instead of the sum of the absolute values, it uses the sum of the squares of the coefficients. This penalty helps to shrink the coefficients, but unlike Lasso, it does not set any coefficients to exactly zero. Ridge regression is useful for handling multicollinearity and improving model generalization.

In short, this is an extension to linear regression that introduces a regularization term to reduce model complexity and help prevent overfitting. It is working on value/magnitude of coefficient is almost equal to zero.

$$\text{Cost function} = \text{loss} + \lambda \sum ||w||^2$$

Loss – sum of squared residual, λ – penalty, w – slope of curve