Regularization

Created by www.linkedin.com/in/yogitadarade

What is Regularization?

Linear regression algorithm works by selecting coefficients for each independent variable that minimizes a loss function. However, if the coefficients are large, they can lead to over-fitting on the training dataset, and such a model will not generalize well on the unseen test data.This is where regularization helps. Regularization is the process which regularizes or shrinks the coefficients towards zero. In simple words, regularization discourages learning a more complex or flexible model, to prevent overfitting.

How Does Regularization Work?

Regularization basically adds the penalty as model complexity increases.

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2$$

Linear Regression cost function

The cost function above is called the residual sum of squares (RSS).Based on our training data, it will adjust the weights (coefficients). If our dataset is noisy, it will face overfitting problems and estimated coefficients won't generalize on the unseen data. This technique regularizes these learned estimates towards zero by penalizing the magnitude of coefficients.

Main Regularization Techniques

- Ridge Regression (L2 Regularization)
- Lasso Regression (L1 Regularizaion)
- Elastic Net Regression

Ridge Regression(L2)

Ridge regression adds “squared magnitude” of coefficient as penalty term to the loss function.

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2$$

When λ = 0, the penalty term has no effect, and the estimates produced by ridge regression will be equal to least squares. However, as λ increase, the impact of the shrinkage penalty grows, and the ridge regression coefficient estimates will approach zero.

Lasso Regression(L1)

Lasso adds "absolute values of magnitude of coefficient as penalty term to the loss function

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

Lambda continues to be input parameters which will decide how high the penalties would be for the coefficients.Larger the value , the lesser will the coefficients be .

Created by www.linkedin.com/in/yogitadarade

Elastic Net Regression

Elastic net regression combines the properties of ridge and lasso regression. It works by penalizing the model using both the 1l2-norm1 and the 1l1-norm1.

Elastic Net Formula: Ridge + Lasso

Key difference between Lasso and Ridge.

- Lasso regression tends to make coefficients to absolute zero as compared to Ridge which never sets the value of coefficient to absolute zero . It can be used for feature selection.
- Lasso does not work well with multicollinearity.Lasso might randomly choose one of the multicollinear variables without understanding the context. Such an action might eliminate relevant independent variables.

Created by www.linkedin.com/in/yogitadarade