Model Building

Bookmark this page

Code

Finally we have arrived at most interesting stage of the whole process — predictive modeling. We will start off with the simpler models and gradually move on to more sophisticated models. We will start with the simpler linear models and then move over to more complex models like RandomForest and XGBoost.

We will build the following models in the next sections.

- Linear Regression
- Lasso Regression
- Ridge Regression
- RandomForest
- XGBoost

Evaluation Metrics for Regression

The process of model building is not complete without evaluation of model's performance. That's why we need an evaluation metric to evaluate our model. Since this is a regression problem, we can evaluate our models using any one of the following evaluation metrics:

- **Mean Absolute Error (MAE)** is the mean of the absolute value of the errors:

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **Mean Squared Error (MSE)** is the mean of the squared errors:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- **Root Mean Squared Error (RMSE)** is the square root of the mean of the squared errors:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

At the competition's page, it has been mentioned that our submission data would be evaluated based on the RMSE score. Hence, we will use RMSE as our evaluation metric.

To know more about evaluation metrics, read this article.

© Analytics Vidhya. All rights reserved except where noted. EdX, Open edX and their respective logos are trademarks or registered trademarks of edX Inc.

Terms of Use - Privacy Policy - Refund Policy