Team

Rules

Overview

Overview

Description

Evaluation

Tutorials

Frequently Asked Questions

Data

Code

Kaggle Learn

Leaderboard

Discussion

Kaggle Learn offers hands-on courses for most data science topics. These short courses prepare you with the key ideas to build your own projects.

My Submissions

Submit Predictions

The Machine Learning Course will give you everything you need to succeed in this competition and others like it.

Other R Tutorials

Fun with Real Estate Data

Use Rmarkdown to learn advanced regression techniques like random forests and XGBoost

XGBoost with Parameter Tuning

- Implement LASSO regression to avoid multicollinearity
- Includes linear regression, random forest, and XGBoost models as well

Ensemble Modeling: Stack Model Example

- Use "ensembling" to combine the predictions of several models
- Includes GBM (gradient boosting machine), XGBoost, ranger, and neural net using the caret package

A Clear Example of Overfitting

Learn about the dreaded consequences of overfitting data

Other Python Tutorials

Comprehensive Data Exploration with Python

- Understand how variables are distributed and how they interact
- Apply different transformations before training machine learning models

House Prices EDA

- Learn to use visualization techniques to study missing data and distributions
- Includes correlation heatmaps, pairplots, and t-SNE to help inform appropriate inputs to a linear model

A Study on Regression Applied to the Ames Dataset

- Demonstrate effective tactics for feature engineering
- Explore linear regression with different regularization methods including ridge, LASSO, and ElasticNet using scikit-learn

Regularized Linear Models

- Build a basic linear model
- Try more advanced algorithms including XGBoost and neural nets using Keras

This competition runs indefinitely with a rolling leaderboard which invalidates entries after two months.

10,552 11,045 67,238

Teams Competitors Entries

Points This competition does not award ranking points

Tiers This competition does not count towards tiers

Tags

regression

tabular data

rmsle