

Module 8: Feature Engineering and Overfitting

Glossary

Inference

A statement about the relationship between the response and the predictor variables

Overfitting

A modeling error where a function is too closely fit to a limited set of data points resulting in an inability of the model to generalize to other datasets

Pipeline

A scikit-learn functionality that assembles several steps that can be cross-validated together while setting different parameters

Prediction

A conclusive statement that pertains to a future event or occurrence

Test Set

A subset of the initial dataset that is set aside to test the trained model

Training Set

The dataset you use to train an algorithm or machine learning model

Validation Set

A subset of the initial dataset that is used to give an estimate of model skill while tuning the model's hyperparameters