Machine Learning Fundamentals - Introduction

 (https://github.com/learn-co-curriculum/dsc-ml-fundamentals-intro) **(https://github.com/learn-co-curriculum/dsc-ml-fundamentals-intro/issues/new/choose)**

Introduction

Linear regression serves as the bridge between traditional statistics and machine learning in this curriculum. In this section you'll learn how to apply machine learning techniques using the familiar linear regression algorithm.

Statistical Learning Theory

Statistical learning goes beyond the statistical modeling we have covered already. We will continue to build models that model the relationships between independent and dependent variables, but the emphasis will shift from *inference* to *prediction*. In the predictive context we also need to consider model *generalization* and not just how well the model is fit to the training data.

Model Validation

In order to assess how well a predictive model will generalize to unseen data, we need model *validation* techniques. These include a *train-test split* and *cross-validation*.

Bias-Variance Trade-Off

Predictive modeling often involves striking the right balance between *bias* and *variance*, also referred to as the balance between *underfitting* and *overfitting*. Especially as you learn models beyond linear regression, you'll be able to tune the model performance to strike the right balance.

Regularization

Regularization is a technique to help prevent overfitting in predictive modeling. We'll specifically discuss *ridge* and *lasso* regression, which are extensions to linear regression that include penalty terms to help prevent overfitting.

Lasso regression in particular does not have a closed form solution and therefore must be solved using an alternative approach such as gradient descent. It also can be helpful for feature selection purposes.

Summary

Linear regression can be used for prediction as well as inference. This has implications for the modeling techniques required, because the emphasis is not solely on the fit to the training data. In this section we'll go over the theoretical concerns as well as the practical approaches to tackling them.

How do you feel about this lesson?

Have specific feedback?

Tell us here! ⤷ **(https://github.com/learn-co-curriculum/dsc-ml-fundamentals-intro/issues/new/choose)**