

Assignment 5 – Prevent Overfitting

Description:

In statistics, overfitting is "the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably." Overfitting could happen to almost any kind of supervised machine learning method. For this assignment, you are going to use L1 Norm, L2 Norm, or the combination of L1 and L2 Norms as the regularization method to prevent the polynomial regression overfit the training data.

Purpose:

- Build polynomial regression model.
- Compare different regularization method.
- Sending data to Google Colab.

Directions:

For this assignment, you need to first build and overfit a polynomial regression. Then, you need to use different regularization methods to prevent the overfitting of that model. Below is a detailed instruction of what you may need to do.

- Dataset Preparation
 - § A dataset is given. If you use Google Colab, you need to upload the dataset to your Google Drive, and access the data in Google Colab.
- Polynomial Regression Design and Overfitting
 - o You need to design a polynomial regression model, then, overfit the polynomial model on the given dataset.
 - § You could use the *scikit-learn* build-in functions to fit the polynomial model.
 - § (Optional) You could develop the polynomial regression model from scratch.
 - Bonus points will be given if you build your model from scratch.
- Prevent Overfitting
 - o After you overfitted the polynomial regression model, you may apply L1 norm, L2 norm, and the combination of L1 and L2 norms to prevent overfitting.
 - § You could use *scikit-learn* build-in functions such as
 - *sklearn.linear_model.Lasso* to build a polynomial regression model with L1 norm
 - *sklearn.linear_model.Ridge* to build a polynomial regression model