PGP in AI/ML - Regression Project Part D

Submission Deadline - 23:59 18th July 2019

Total Marks: 14

Objective

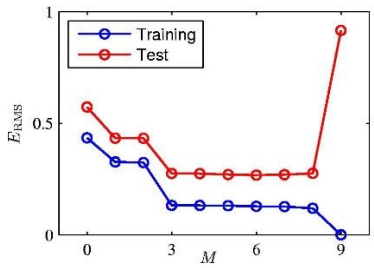
The goal of this part of the project is to build a polynomial regression model for a given dataset and demonstrate **Lasso and Ridge** regressions. The regression models should be implemented using the libraries sklearn and numpy.

Problem Description

You should use the dataset that is coming up with this assignment. Please observe that this data set is a subset of the Father-Son dataset that has been used for previous assignments. There are 2 files, train.csv (20 data points) and test.csv (5 data points).

1. Implement Polynomial regression (degrees 1 to 10)

- Plot train and test RMSE, for each degree. These plots can be graphs where degree of the polynomial is x-axis and RMSE is y-axis. The graphs should be similar to the following graphs (obviously your graphs will be different but they look like as follows):



- Select the best degree of polynomial based on test error (lower the better.)

2. Regularization using Lasso and Ridge

- For degree 10, find out the train and test RMSE for ridge and lasso regression. You may take the regularization parameter (alpha or lambda) to be the default parameter provided by python. Discuss the improvements brought down by ridge, lasso regression in RMSE as compared to the typical regression model built using the polynomial of degree 10.

Polynomial regression implementation

We do not have a direct API for polynomial regression in scikit-learn library. So, we have to treat polynomial regression ($y = w_0 + w_1 x + w_2 x^2 + \ldots. + w_D x^D$) as a multiple regression problem ($y = w_0 + w_1 x_1 + w_2 x_2 + \ldots. + w_D x_D$) with $x1 = x$, $x_2 = x^2$, $x_3 = x^3$ …… $x_D = x^D$ to build polynomial regression model of degree 'D'.

1. Create an object of sklearn.preprocessing.PolynomialFeatures class.
2. Parameter degree of the class represents the degree of polynomial that we are trying to fit the data.
3. We convert single feature X into $[\mathbf{1}, \boldsymbol{X}, \boldsymbol{X^2}, \boldsymbol{X^3}, \boldsymbol{X^4}, \ldots., \boldsymbol{X^D}]$, where D represents the parameter degree of the class by applying fit_transform() method on X(features) and store it in an object modified_X.
4. Generate a linear regression model similar to Part-A with passing modified_X instead of X in the fit() method for training.

Lasso/Ridge implementation

Instead of using sklearn.linear_model.LinearRegression for building simple linear regression models, you have to make of sklearn.linear_model.Lasso, and sklearn.linear model.Ridge to build Lasso and Ridge regression models respectively.

Submission Details

1. Polynomial Regression Code - id_poly_part_d.py. Plot train and test RMSE scores. (8 M)

2. Lasso Code - id_lasso_part_d.py. Find out the train and test RMSE. (3M)

3. Ridge Code - id_ridge_part_d.py. Find out the train and test RMSE. (3M)

Contacts

You should put up questions in the discussion forum of the corresponding assignment folder only. Please put all your queries to the following TAs in Canvas but not to the instructor.

- Kartheek f20160015@hyderabad.bits-pilani.ac.in
- Shristy f20160688@hyderabad.bits-pilani.ac.in
- Agam f20160484@hyderabad.bits-pilani.ac.in
- Himanshu himanshuchawla437@gmail.com
- Mohit mohitrathore15@gmail.com