Machine Learning

1st Assignment - Shahid Beheshti University

February 26, 2023

Due date: April 3rd

You are required to write a detailed report for implementation tasks. **

1. Can gradient descent get stuck in a local minimum when training a logistic regression model? Why?

2. Suppose you are using polynomial regression. You plot the learning curves and you notice that there is a large gap between the training error and the validation error. What is happening? What are three ways to solve this?

3. Suppose you are using ridge regression and you notice that the training error and the validation error are almost equal and fairly high. Would you say that the model suffers from high bias or high variance? Should you increase the regularization hyperparameter α or reduce it?

4. Why would you want to use:

- Ridge regression instead of plain linear regression (i.e., without any regularization)?
- Lasso instead of ridge regression?
- Elastic net instead of lasso regression?

5. Implement Linear Regression with **Mean Absolute Error** as the cost function from scratch. Compare your results with the Linear Regression module of **Scikit-Learn**.

6. Implement Linear Regression using the normal equation as the training algorithm from scratch.

7. Compare bootstrapping with cross-validation. In which conditions we should use bootstrapping?

8. Explain nested cross-validation and 5x2 cross-validation in detail and when we should use them.*(Extra Point)*

9. How can we compare different models using statistical significance tests? *(Extra Point)*

10. Implement Forward and Backward Feature selection algorithms from scratch with MSE as the metric.

11. Suppose the features in your training set have very different scales. Which algorithms *(Gradient Descent, Normal Equation, SVD)* might suffer from this, and how? What can you do about it?

12. In this part, you are going to work with the **News Popularity Prediction** dataset. You will implement a regression model using the **Scikit-Learn** package to predict the *popularity of new articles* *(the number of times they will be shared online)* based on about 60 features. You are expected:

- Perform exploratory data analysis on the dataset.
- Propose 5 different hypothesis tests related to the dataset. At least use 3 different tests.
- Try Ridge and Lasso regression.
- Use various scaling methods and report their effects.
- Add polynomial features and report their effect.
- Try using GridSearchCV with RandomizedSearchCV to tune your model's hyperparameters. *(Extra Point)*
- Apply the feature selection methods that you have implemented in the above sections.
- Get familiar with and implement the following loss functions from scratch and utilize them with a Linear Regression model and discuss their effect on the performance of the model. *(Extra Point)*
  - Absolute Error
  - Epsilon-sensetive error
  - Huber

13. Implement batch gradient descent with early stopping for *softmax regression* from scratch. Use it on a classification task on the Penguins dataset. *(Extra Point)*

Good Luck!