Assignment 1

Neural Network and Fuzzy Logic

Instructor: Dr. Rajesh Kumar Tripathy

(Assignment must be done using Python with Google Colab framework)

Submission date: (October 5, 2021, hard deadline)

Q1. Implement linear regression (LR) approach using batch gradient descent (BGD), stochastic gradient descent (SGD), and mini-batch gradient descent (MBGD) algorithms. Show the cost-function vs. epoch plots for LR with BGD, LR with SGD, and LR with MBGD models. Show the contour plots for cost function vs. w1 vs. w2 evaluated using LR with BGD, LR with SGD, and LR with MBGD models. For Q1, the **data-q1.xlsx** file must be used. The data q1.xlsx file contains two inputs and one output. You can consider w1 and w2 are the weight values of features.

Q2. Implement linear regression with the L2-norm regularization (Ridge regression) approach using BGD, SGD, and MBGD algorithms. The ridge regression model weight parameters must be evaluated from the training data. After evaluating the weight parameters, evaluate the predicted output for the parameters, evaluate the predicted output for the test feature vectors. For Q2, the **data_q2_q3.xlsx** file must be used. Evaluate the mean square error (MSE), mean absolute error (MAE), and correlation coefficient (CC) by comparing the actual test output and predicted test output for ridge regression models with BGD, SGD, and MBGD algorithms. You can use grid search to evaluate the optimal parameters of the model. You can divide the dataset into training, validation, and testing using hold-out cross-validation (70% (training), 10% (validation), and 20% (testing)).

Q3. Repeat question no. Q2 using least angle regression models with BGD, SGD, and MBGD algorithms. Evaluate MSE, MAE, and CC values for the test data. You can use grid search to evaluate the optimal parameters of the models. You can divide the dataset into training, validation, and testing using hold-out cross-validation (70% (training), 10% (validation), and 20% (testing)). For Q3, the **data_q2_q3.xlsx** file must

be used.

Q4. Implement logistic regression (LOR), LOR with L2-norm regularization, and LOR with L1-norm regularization models using BGD, SGD, and MBGD algorithms. The dataset in data_q4_q5.xlsx contains 30 features and one output. The class label '*M*' stands for malignant, and '*B*' is the Benign class. You must use hold-out cross-validation ((CV) with 70% as training, 10% as validation and 20% as testing) to evaluate training, validation, and testing instances for each model. Evaluate the performance of each model using accuracy, sensitivity, and specificity measures.

Q5.Repeat the Q4 using a 5-fold CV-based selection of training and test instances for each model. Evaluate the accuracy, sensitivity, and specificity values of LoR, LoR+L2-norm regularization, LoR+L1-norm regularization models using BGD, SGD, and MBGD algorithms. You must use the dataset data_q4_q5.xlsx for this question.

Q6. Implement multiclass LOR, multiclass LOR with L2-norm regularization, and multiclass LOR with L1-norm regularization models using BGD, SGD, and MBGD algorithms. The multiclass extension of the LOR models must be done using One vs. one and one vs. All coding algorithms. The dataset in data_q6_q7.txt contains 7 features and one output. The output is classified as class 1, class2, or class 3. You must use hold-out cross-validation ((CV) with 70% as training, 10% as validation and 20% as testing) for the evaluation of training, validation, and testing instances for each model. Evaluate the performance of each model using individual accuracy and overall accuracy measures.

Q7. Repeat Q7 using a 5-fold CV-based selection of training and test instances for each model. Evaluate the accuracy, sensitivity, and specificity values of multiclass LoR, multiclass LoR+L2-norm regularization, multiclass LoR+L1-norm regularization models using BGD, SGD, and MBGD algorithms. Evaluate the performance of each model using individual accuracy and overall accuracy measures. You must use the

dataset data_q6_q7.txt for this question.

Q8. Use the likelihood ratio test (LRT) for the binary classification using the dataset ("data_q4_q5.xlsx"). You must use a 5-fold CV-based selection of training and test instances to evaluate the LRT classifier. Evaluate the accuracy, sensitivity, and specificity values for the binary classifier.

Q9. Implement the Maximum a posteriori (MAP) decision rule for the multiclass classification tasks. You must use a 5-fold CV-based selection of training and test instances for the MAP classifier. You must use the dataset data_q6_q7.txt for this question. Evaluate individual accuracy and overall accuracy of MAP classifier.

Q10. Implement the Maximum likelihood (ML) decision rule for the multiclass classification task. Use the hold-out cross-validation approach (70% training and 30% testing) for the selection of training and test instances of the ML classifier. You must use the dataset data_q6_q7.txt for this question. Evaluate individual accuracy and overall accuracy of ML classifier.

Q11. Implement the K-means clustering-based unsupervised learning algorithm for the dataset ("dataq11.xlsx"). Plot the estimated class labels vs individual features. Use the number of clusters as K=20.

More instructions

1. You can use pandas, NumPy, Matplotlib, and Scipy (only for the matrix part) 2. Keras, sci-kit-learn, TensorFlow, etc. are not allowed. If you use these packages, then questions will

not be checked.