Mouhammadou Dabo

March 18, 2021

CS 1675: Intro to Machine Learning

Professor Milos Hauskrecht

Problem Assignment 5

Problem 1. Support vector machines

Part a.

w (weights)

$$= \begin{bmatrix} 0.0445 \\ 0.0116 \\ -0.0050 \\ 0.0003 \\ -0.0002 \\ 0.0264 \\ 0.3524 \\ 0.0063 \end{bmatrix}$$

b (bias)

$$= -2.7182$$

Part b.

apply_svlm.m

Part c.

Confusion matrix for train set

$$\begin{bmatrix} 0.5547 & 0.0742 \\ 0.1577 & 0.2134 \end{bmatrix}$$

Training misclassification error

$$E = FP + FN$$

$$E = 0.0742 + 0.1577 = 0.2319$$

Sensitivity of the model on the train set

$$SN = \frac{TP}{TP + FN} = \frac{0.2134}{0.2134 + 0.1577} = 0.5750$$

Specificity of the model on the train set

$$SP = \frac{TN}{TN + FP} = \frac{0.5547}{0.5547 + 0.0742} = 0.8820$$

Confusion matrix for test set

$$\begin{bmatrix} 0.6201 & 0.0830 \\ 0.1135 & 0.1834 \end{bmatrix}$$

Testing misclassification error

$$E = FP + FN$$

$$E = 0.0830 + 0.1135 = 0.1965$$

Sensitivity of the model on the test set

$$SN = \frac{TP}{TP + FN} = \frac{0.1834}{0.1834 + 0.1135} = 0.6176$$

Specificity of the model on the test set

$$SP = \frac{TN}{TN + FP} = \frac{0.6201}{0.6201 + 0.0830} = 0.8820$$

Part d.

Comparing the results I received after using support vector machines (SVMs) with the results for the logistic regression model (Assignment 5), this model performed way better. The results from using SVMs were comparable to the results I received after optimizing the parameters of the linear regression model, with a training error of around 0.23 and testing error around 0.20; these results for the linear regression model were discovered after experimenting with the number of epochs (30,000) and learning rates (0.001) along with consistent weights.

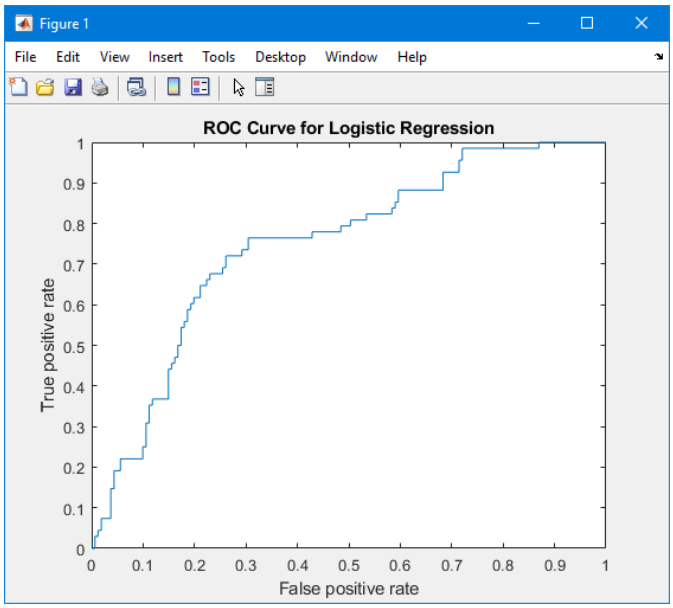
Problem 2. ROC analysis

Part a. N/A

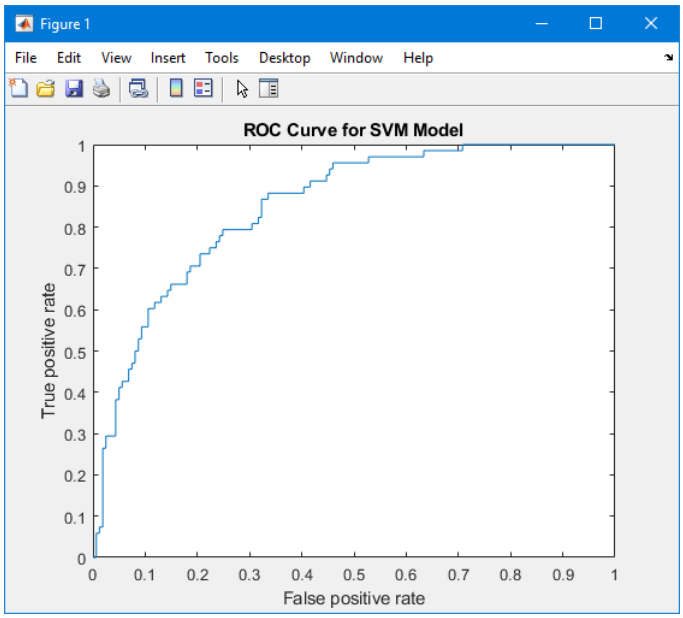
Part b. N/A

Part c.

ROC Curve for Logistic Regression



ROC Curve for SVM Model



AUROC for Logistic Regression

$$= 0.74872$$

AUROC for SVM Model

$$= 0.84974$$

By comparing the ROC curves and AUC statistics for the two models, the SVM model performed better than the logistic regression model. Considering the closer the area under the ROC is better when closer to 1 and worst when closer to 0.5, the AUROC for the SVM model is higher than that of the logistic regression model making a better model. The ROC curve of the SVM model also resembles more of a rectangular shape than that of the logistic regression model.

Problem 3. Deep learning toolbox in Matlab

Part a. N/A

Part b.

NNs with one hidden layer, 2000 epochs

Hidden Units	**2**	**3**	**5**	**10**
Training Error	0.2356	0.2449	0.2263	0.2152
Testing Error	0.2183	0.2096	0.2183	0.2140

After comparing the mean misclassification errors for the training and testing data from the logistic regression model and the multilayer neural network, the multilayer neural network performed better even when only 2 hidden units were present. The training and testing errors from the logistic regression model were 0.3061 and 0.2707 respectively. Experimenting with different hidden units for the multilayer neural network showed that increasing the number of hidden units provided better errors, but with diminishing returns; by testing and comparing the number of hidden units, you can optimize the model to produce better training and testing errors.