CS573_assignment3_XiaofengOu

Xiaofeng Ou

March 2019

1 Preprocessing

Output:

```
Mapped vector for female in column gender: [1.]
Mapped vector for Black/African American in column race: [0. 1. 0. 0.]
Mapped vector for Other in column race_o: [0. 0. 0. 0.]
Mapped vector for economics in column field:
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
```

2 Logistic Regression and Linear SVM

Logistic Regression:

```
Training Accuracy LR:  0.65
Testing Accuracy LR:  0.65
```

SVM:

```
Training Accuracy SVM:  0.56
Testing Accuracy SVM:  0.55
```

3 Learning Curves and Performance Comparison

Hypothesis :

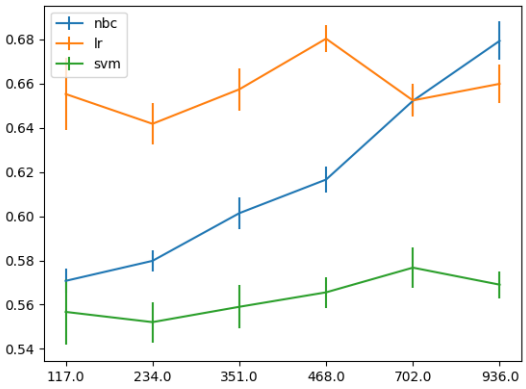

Figure 1: Models Accuracy

- H_0: On this data set, the performance of logistic regression and SVM are significantly the same.
- H_1: On this data set, the performance of logistic regression and SVM are significantly not the same.

Hypothesis Testing: I used t-test. The p value of the t-test is

```
p value for the t test:  2.6943077293774723e-08
```

. We should reject the hypothesis H_0. That is, on this data set, the performance of logistic regression and SVM are significantly not the same.