CS57300: Homework 3

1. ***Assess whether choice of model improves performance.***

(a) Plot the learning curves for the three models (in the same plot), including error bars that indicate ±1 standard error, from the evaluation based on incremental CV as described above.

<u>Ans.</u>

The following are the Zero-One-Losses for train size vs. K-fold number:

For Logistic Regression:

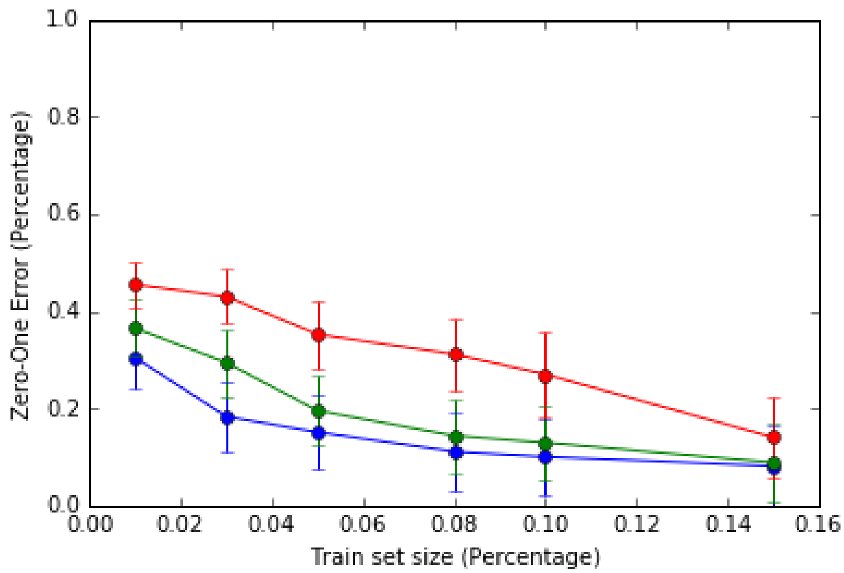
	1	2	3	4	5	6	7	8	9	10	Mean	Std Error
0.01	0.245	0.330	0.295	0.275	0.165	0.330	0.390	0.440	0.420	0.270	0.430	0.064
0.03	0.140	0.225	0.180	0.140	0.150	0.195	0.300	0.115	0.175	0.190	0.318	0.073
0.05	0.070	0.130	0.150	0.100	0.245	0.185	0.130	0.195	0.090	0.155	0.288	0.076
0.08	0.110	0.140	0.090	0.055	0.105	0.115	0.100	0.085	0.100	0.105	0.250	0.079
0.1	0.085	0.090	0.090	0.075	0.055	0.095	0.120	0.075	0.060	0.140	0.240	0.080
0.15	0.055	0.070	0.045	0.065	0.020	0.100	0.100	0.065	0.085	0.070	0.223	0.082

For Support Vector Machines:

	1	2	3	4	5	6	7	8	9	10	Mean	Std Error
0.01	0.320	0.345	0.450	0.280	0.345	0.275	0.515	0.405	0.425	0.475	0.486	0.058
0.03	0.235	0.170	0.220	0.180	0.370	0.440	0.440	0.405	0.150	0.430	0.420	0.069
0.05	0.270	0.280	0.235	0.150	0.280	0.100	0.125	0.100	0.195	0.210	0.329	0.073
0.08	0.155	0.175	0.215	0.100	0.105	0.150	0.130	0.095	0.095	0.150	0.281	0.076
0.1	0.140	0.080	0.140	0.120	0.075	0.105	0.170	0.095	0.140	0.145	0.268	0.077
0.15	0.085	0.070	0.085	0.065	0.055	0.070	0.095	0.060	0.080	0.090	0.230	0.081

For Naïve Bayes Classifier:

	1	2	3	4	5	6	7	8	9	10	Mean	Std Error
0.01	0.505	0.465	0.455	0.550	0.445	0.445	0.550	0.495	0.425	0.505	0.570	0.047
0.03	0.505	0.505	0.225	0.550	0.520	0.540	0.550	0.345	0.400	0.415	0.546	0.056
0.05	0.505	0.505	0.455	0.135	0.430	0.360	0.135	0.345	0.575	0.240	0.474	0.070
0.08	0.315	0.190	0.455	0.120	0.310	0.115	0.550	0.500	0.165	0.505	0.435	0.075
0.1	0.125	0.505	0.185	0.550	0.130	0.070	0.110	0.500	0.080	0.505	0.397	0.085
0.15	0.065	0.055	0.090	0.075	0.520	0.070	0.080	0.120	0.125	0.125	0.277	0.085



As can be seen, the **NBC(red)** classifier is the least accurate for small test-train ratios as compared to the **SVM(green)** and the most accurate for such small training data size is the **Logistic Regression (blue)**. Here, the **SVM(green)** and the **Logistic Regression (blue)** have been noticed to be very close, often one being more efficient than the error, as can be seen form the standard errors.

(b) Formulate a hypothesis about the performance difference between at least two of the models.

Ans.

We can formulate three hypotheses:

Hypothesis 1:

H_0: The average performance of the NBC classifier is the same as that of the SVM classifier.

H_1: The average performance of the NBC classifier is not same as that of the SVM classifier.

Hypothesis 2:

H_0: The average performance of the SVM classifier is the same as that of the LR classifier.

H_1: The average performance of the SVM classifier is not same as that of the LR classifier.

Hypothesis 3:

H_0: The average performance of the NBC classifier is the same as that of the LR classifier.

H_1: The average performance of the NBC classifier is not same as that of the LR classifier.

(c) Discuss whether the observed data support the hypothesis (i.e., are the observed differences significant).

Ans.

Hypothesis 1:

As can be seen in the graph, the performance of the NBC classifier starts off as moderately low-performance compared to the SVM as can be seen by the significant difference in the means for the various test sizes, and that the standard errors do not overlap between sizes of 0.03 to 0.01 (including).

From this, we can conclude that since most points have significantly different means, and since most standard error regions do not overlap, we reject null hypothesis and say that they are in fact different.

Hypothesis 2:

As can be seen in the graph, the performance of the SVM classifier starts off as moderately low-performance compared to the LR classifier as can be seen by the significant difference in the means for the various test sizes, but the standard errors are always overlapping.

From this, we can conclude that since most points have significantly different means, and even though most standard error regions do overlap, we reject null hypothesis and say that they are in fact different.

Hypothesis 3:

As can be seen in the graph, the performance of the NBC classifier starts off as highly low-performance compared to the LR as can be seen by the significant difference in the means for the various test sizes, and that the standard errors do not overlap significantly.

From this, we can conclude that since most points have significantly different means, and since most standard error regions do not overlap, we reject null hypothesis and say that they are in fact different.

2. **Assess whether feature construction affects performance.**

(a) **Plot the learning curves for the three models (in the same plot), including error bars that indicate ±1 standard error, from the evaluation based on incremental CV as described above.**

Ans.

The following are the Zero-One-Losses for train size vs. K-fold number:

For Logistic Regression:

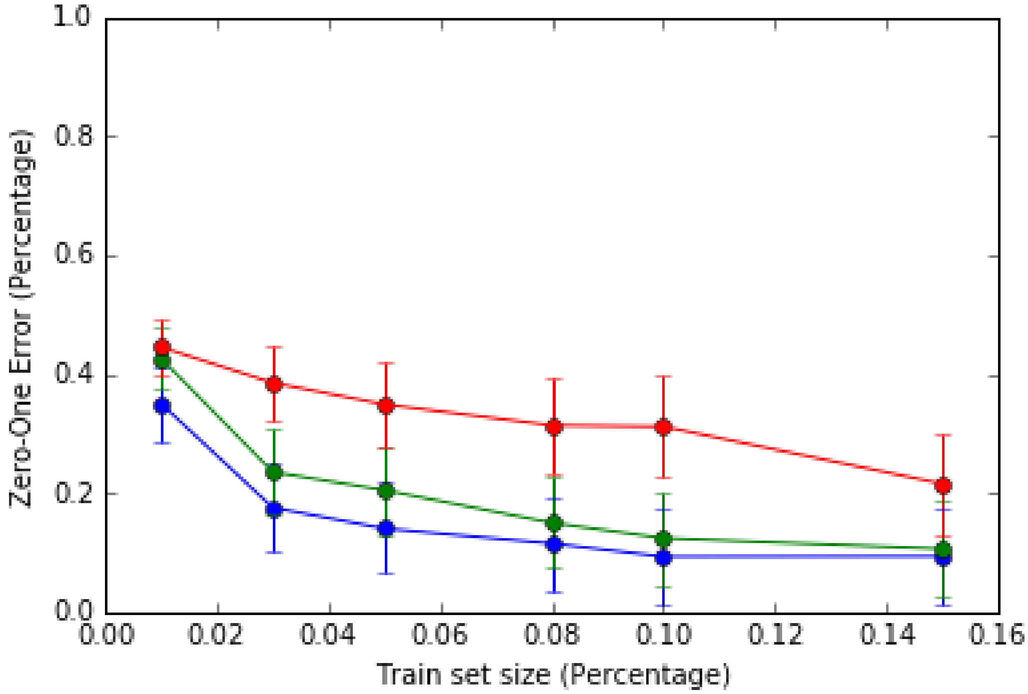
	1	2	3	4	5	6	7	8	9	10	Mean	Standard Error
0.01	0.410	0.385	0.250	0.230	0.235	0.235	0.315	0.345	0.375	0.295	0.423	0.063
0.03	0.140	0.175	0.150	0.150	0.125	0.155	0.135	0.155	0.135	0.225	0.295	0.074
0.05	0.080	0.100	0.125	0.110	0.095	0.150	0.110	0.190	0.070	0.100	0.261	0.078
0.08	0.080	0.095	0.105	0.090	0.070	0.080	0.105	0.080	0.105	0.080	0.241	0.080
0.1	0.085	0.070	0.070	0.080	0.090	0.080	0.085	0.095	0.065	0.070	0.233	0.080
0.15	0.040	0.095	0.040	0.060	0.055	0.095	0.100	0.075	0.080	0.070	0.226	0.081

For Support Vector Machines:

	1	2	3	4	5	6	7	8	9	10	Mean	Standard Error
0.01	0.480	0.385	0.325	0.440	0.270	0.235	0.435	0.360	0.505	0.235	0.472	0.062
0.03	0.190	0.220	0.150	0.190	0.390	0.425	0.205	0.125	0.145	0.215	0.355	0.073
0.05	0.095	0.225	0.115	0.145	0.090	0.210	0.150	0.245	0.180	0.265	0.310	0.074
0.08	0.175	0.130	0.120	0.125	0.085	0.175	0.140	0.080	0.110	0.120	0.272	0.077
0.1	0.105	0.160	0.085	0.100	0.110	0.125	0.150	0.115	0.065	0.115	0.261	0.078
0.15	0.060	0.135	0.050	0.060	0.080	0.115	0.090	0.115	0.100	0.070	0.240	0.080

For Naïve Bayes Classifier:

	1	2	3	4	5	6	7	8	9	10	Mean	Standard Error
0.01	0.520	0.405	0.530	0.575	0.475	0.495	0.465	0.470	0.510	0.520	0.497	0.014
0.03	0.410	0.570	0.330	0.420	0.475	0.125	0.400	0.370	0.165	0.590	0.385	0.045
0.05	0.150	0.320	0.335	0.115	0.290	0.410	0.430	0.125	0.145	0.590	0.291	0.047
0.08	0.560	0.140	0.175	0.135	0.265	0.255	0.315	0.505	0.090	0.160	0.290	0.048
0.1	0.350	0.070	0.420	0.165	0.370	0.145	0.250	0.200	0.370	0.380	0.272	0.036
0.15	0.125	0.065	0.110	0.545	0.120	0.165	0.100	0.140	0.200	0.165	0.174	0.041



As can be seen, the **NBC(red)** classifier is the least accurate for small test-train ratios as compared to the **SVM(green)** and the most accurate for such small training data size is the **Logistic Regression (blue)**. Here, the **SVM(green)** and the **Logistic Regression (blue)** have been noticed to be very close, often one being more efficient than the error, as can be seen form the standard errors.

It can however be noticed that in general, the **NBC(red)** classifier here has a worse prediction accuracy than in the case of 0-1 binary feature construction **NBC(red)**.

(b) **Formulate a hypothesis about the performance difference you observe for at least one model (comparing results from this experiment and (1)).**

Ans.

H_0: The average performance of the LR (new) is the same as that of the LR (old).

H_1: The average performance of the LR (new) is not the same as that of the LR (old).

(c) **Discuss whether the observed data support the hypothesis (i.e., are the observed differences significant).**

Ans. As can be seen from the plots, from first glance the two-classifier's accuracy seems to be the same. To confirm this, a t-test was conducted with the following results:

	Variable 1	*Variable 2*
Mean	0.279833	0.279722
Variance	0.005536	0.005551
Observations	6	6
Pooled Variance	0.005543	
Hypothesized Mean Difference	0	
df	10	
t Stat	0.002585	
P(T<=t) one-tail	0.498994	
t Critical one-tail	1.812461	
P(T<=t) two-tail	0.997988	
t Critical two-tail	2.228139	

As can be seen, since the p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis.