CS440 FINAL PROJECT REPORT

Face and Digit Classification

Authors (RUID)
Ahmed Ishaque (199006840)
Andy Cheung (202002593)
Gabriel Weinbrenner (200004871)

1 Introduction

2 Implementation

2.1 Naive Bayes

This implementation utilizes probabilities to classify information. There are three counters that measure different variables. The first is used for the labels P(Y). This value is estimated by dividing the label y by the total number of training instances evaluated through the training data. The resulting equation is $P(Y) = c(y)/n$. The second counter is a dictionary that contains 0's and 1's to represent black and white features. This binary system is used for conditional probability. The third counter counts the number of instances a specific counter appears. As the first loop iterates through training data, the third counter is increased for each feature that appears, and the first counter counts labels and classifies 0's and 1's. The first counter is normalized after the looping, and the conditional probabilities are smoothed. In order to avoid features with values of 0's, the probabilities are increased by a number to each feature. In the end, the conditional probabilities are normalized.

2.2 Perceptron

The implementation of the Perceptron procedure has been easier, because it is based on a weighted system and not probabilities. This weighted formula compares y' to the true label y. If the comparison of both variables show that they are equal, then it results in a correct classification. Otherwise, the Perceptron procedure guesses y'. The global counter is associated with each label, and a nested loop is ran where the outer loop is the amount of iterations that would be ran. This inner loop filters through the training data. The nested loop is meant to keep track of score for each possible label, which is a counter for pixels values. The resulting pixels values is used to update the weighting, and lets the program know either the previous or current score is higher.

2.3 MIRA

MIRA classifies information by examining data through each separate instance. A global counter updates weights for each label, but only occurs when a classification is incorrect. The value Tau is applied to the formula referenced on Berkeley's website. The minimum of Tau, and the maximum possible value, C, is measured to update the weight vectors, which strengthens the algorithm.

3 Results

3.1 Faces

Faces

Naive Bayes

Percent of training data	Time (seconds)	Accuracy (%)	Standard Dev of Accuracy
10	6.23	75.73	2.65
20	7.15	78.93	5.09
30	7.99	84.0	3.13
40	8.34	87.07	1.92
50	8.34	86.67	1.25
60	8.57	88.8	1.85
70	9.00	89.2	0.87
80	9.74	88.8	1.59
90	9.73	89.2	0.29
100	10.0	90	0

MIRA (Faces)

Percent of training data	Time (seconds)	Accuracy	Standard Dev of Accuracy
10	6.56	71.2%	5.28
20	9.19	76.8%	2.56
30	11.62	77.6%	2.69
40	13.7	78.8%	4.51
50	16.29	82.13%	4.98
60	18.32	83.6%	3.99
70	14.03	84.93%	1.12
80	15.55	84.0%	3.68
90	17.02	85.47%	2.02
100	18.98	86.4%	2.77

Perceptron

Percent of training data	Time (seconds)	Accuracy (%)	Standard Dev of Accuracy
10	10.25	84.4	4.51
20	11.84	89.47	2.47
30	13.51	91.33	1.25
40	15.32	91.87	0.29
50	17.36	92	0
60	19.51	92	0
70	21.38	92	0
80	22.79	92	0
90	24.66	92	0
100	29.92	92	0

3.2 Digits

Naive Bayes

Percent of training data	Time (seconds)	Accuracy (%)	Standard Dev of Accuracy
10	5.67	72.64	2.33
20	5.87	74.98	1.29
30	5.89	75.98	0.295
40	6.14	75.92	0.46
50	6.19	76.06	0.71
60	6.83	76.42	0.63
70	7.12	76.06	0.53
80	6.76	76.16	0.49
90	6.71	76.52	0.217
100	6.87	76.6	0

Perceptron

Percent of training data	Time (seconds)	Accuracy (%)	Standard Dev of Accuracy
10	32.88	76.22	1.67
20	39.70	79.46	1.70
30	44.82	81.22	1.80
40	51.33	80.82	0.923
50	56.32	81.9	0.469
60	63.49	82.3	0.578
70	68.57	81.28	0.531
80	76.62	82	0.308
90	82.34	81.28	0.572
100	90.22	81.02	0.4919

MIRA

Percent of training data	Time (seconds)	Accuracy (%)	Standard Dev of Accuracy
10	24.29	68.64	9.85
20	39.02	73.6	3.54
30	53.32	77.82	2.17
40	71.46	76.18	3.03
50	83.71	77.78	2.19
60	96.17	77.08	2.08
70	110.66	79.04	2.15
80	128.95	76.4	4.24
90	145.04	78.1	3.10
100	161.26	75.3	3.87