Algebraic Classifiers: Supplementary Material

Michael Izbicki

May 7, 2013

This supplemental material contains more details about the experiments we ran, as well as a brief description of the method of moments and how to give algebraic structure to an estimated distribution.

1 Experiments

For the experiments in Figure 3 and 7 of the main paper, we selected 30 different classification data sets from the UCI machine learning repository to perform our experiments. They were selected because they come from a number of different domains and have widely varying characteristics. They are:

data set file	number of attributes	number of class labels	number of data points
abalone.data	8	3	4177
anneal.data	38	5	798
audiology.standardized.data	70	24	200
balance-scale.data	4	3	625
breast-cancer-wisconsin.data	10	2	699
cmc.data	9	3	1473
crx.data	15	2	690
german.data	20	2	1000
glass.data	10	6	214
haberman.data	3	2	306
house-votes-84.data	16	2	435
hypothyroid.data	25	2	3163
ionosphere.data	34	2	351
iris.data	4	3	150
movement_libras.data	90	15	360
new-thyroid.data	5	3	215
pima-indians-diabetes.data	8	2	768
sat.trn	36	6	4435
segmentation.data	19	7	210
segment.dat	19	7	2310
sick-euthyroid.data	25	2	3163
sonar.all-data	60	2	208
soybean-large.data	35	19	307
spambase.data	57	2	4601
tae.data	5	3	151
tic-tac-toe.data	9	2	958
transfusion.data.modified	6	2	748
Wall-following-sensor_readings_24.data	24	4	5456
wpbc.data	34	23	198
zoo.data	17	7	101

All experiments were conducted using 10-fold cross-validation, with the resulting means displayed in the plots. For the homomorphic learners, we could easily have performed more folds; however, we could not have done this for the non-homomorphic learners tested against. For consistency, we used this relatively low number for all trials.

We selected standard parameters for all of our models. When testing the Bayesian classifier and HomStump, we used the normal distribution for continuous data and categorical distribution for discrete data. For AdaBoost, we set the maximum number of boosting rounds to 25. This is a small number of rounds to use for these simple classifiers, and it's likely that using more rounds would have led to better performance. Ideally, we would have chosen the optimal number of rounds via cross-validation.

2 Method of Moments

The method of moments estimates a continuous probability distribution from a list of samples. It is a group learner and can be used as the base distribution for the Bayesian classifier. Given an n-element input vector $(x_1, x_2, ...x_n)$ for training, the formula for the jth moment is:

$$m_j = \sum_{i=1}^{n} x_i^j$$

The model for the distribution d estimated with k moments is the tuple:

$$d = (m_0, m_1, m_2, ..., m_k)$$

The binary operation $d_0 = d_1 \diamond d_2$ is defined as:

$$m_{0,i} = m_{1,i} + m_{2,i}$$

where $m_{j,i}$ is the ith moment of distribution j. The empty element is:

$$\epsilon = (0, 0, 0, ..., 0)$$

And the inverse is:

$$d^{-1} = (-m_0, -m_1, -m_2, ..., -m_k)$$

Estimating a Normal distribution is a special case of this method because the sufficient statistics are the 0^{th}, 1^{st}, and 2^{nd} moments.