CMPS 142: Homework Assignment 2

Raymond Colebaugh - 1377877
Jeffrey Petersen - 1329242
Peter Czupil - 1317993

April 28, 2015

1. (a)

Algorithm	Correct	Incorrect	MAE	RMS	RAE (%)	RSE (%)
Nearest Neighbor (IB1)	768	0	0	0	0	0
Naive Bayes	586	182	0.2811	0.4133	61.8486	86.7082
Logistic Regression	601	167	0.3063	0.3908	67.3928	81.9907

Nearest Neighbor is the most accurate algorithm for this dataset. This is perhaps due to the fact that the standard deviation for this particular data set is pretty high which allows for a decision boundary with ample room on either side.

(b) Running the Weka logistic regression algortihm on the supplied dataset produced the weight vector:

$$A = \begin{bmatrix} -0.1232 \\ -0.0352 \\ 0.0133 \\ -0.0006 \\ 0.0012 \\ -0.0897 \\ -0.9452 \\ -0.0149 \end{bmatrix}$$

Our bias turned out to be equal to 8.40. Therefore the equation for our decision boundary is:

$$\Sigma_i \; a_i * x_i \; + \; 8.40 \; = \; 0$$

By plugging the $x^{(i)}$'s whose prob. distribution is close to 0.50/0.50 into our decision boundary equation, we have found that indeed, the points lie very close it.

(c)

Algorithm	Correct	Incorrect	MAE	RMS	RAE (%)	RSE (%)
Nearest Neighbor (IB1)	539	229	0.2982	0.5461	65.6046	114.5627
Naive Bayes	586	182	0.2841	0.4168	62.5028	87.4349
Logistic Regression	593	175	0.3094	0.3954	68.0819	82.9651

(d) After applying the normalization filter to the dataset, the attributes now have real numbered values ranging from 0 to 1. Running the 10-fold cross validation yields the same accuracies as well because although the attributes now have lesser values, the cumulative value has also decreased correspondingly. The weight vector changed pretty radically however, before normalization the absolute values of the weight vector were very small (between 0 and 1). After the normalization, the absolute values became much greater (anywhere from 2 to 8 on average). This is because with classification, the prediction rule depends on how close a value is to a 0 or a 1. Before the normalization, the attribute values were large and thus must be multiplied by a small number to get a number between 0 and 1. After normalization the roles have reversed thus the weights are large.

(e) When changing the ridge parameter to 0, the value of the weights don't seem to change, neither does the accuracy. When changing the ridge parameter to 0.3 however, the values of the weights decreases slighty while the accuracy stays the same.

(f) We would expect 3NN or 5NN to be more accurate because the feature set is pretty large thus many combinations of features that lead to a certain label exist. Therefore comparing an instance to more than one of its nearest neighbors reveals more combinations than comparing an instance to just one nearest neighbor. The accuracy of the prediction when running the ibk algorithm in Weka supports this hypothesis. 3NN yields an accuracy of %72.6563 and 5NN gives %73.1771.

(g) We would expect the results to be less accurate when running the classifiers because there is now a lot more random noise. This is not the case with logisitc regression, the accuracy actually stays the same. IB1 had a slight drop in accuracy (around %1) with folding (using the entire training set as the test set kept an accuracy of %100 for obvious reasons) and naive bayes had a much larger drop in accuracy (around %8).

(h) When we ran nearest neighbor on the normalized (features) modified training set, 10 fold cross-validation was more inaccurate than nearest neighbor learning from the training sets of the previous parts. The accuracy when using the entire training set as the test set was %100 however. Naive bayes was more accurate than the previous two training sets and logistic regression was slightly more accurate than the previous training sets as well.

2. (a) The outcome space is the set of possible combinations of whether each of the two children were male or female. Given in the notation of $(younger, older)$, this leads to the outcome space:

$$(B, G), (G, B), (B, B), (G, G)$$

(b) We want the probability of at least one child being a girl, given that we already know one child is a boy. We reduce the outcome space by removing the possiblity of (G, G). Then the remaining outcomes are $(B, G), (G, B), (B, B)$. This leaves us with a probability of $\frac{2}{3}$.

(c) Given that we know one child is a boy, then we are left with two cases: either it was the first or the second child. In the case that we saw the first child, we can ignore the two last possibilities (G, B) and (G, G). In the case that we saw the second child, we can ignore the second and last cases, (B, G) and (G, G)

	First Child		Second Child		
	B	B	B	B	
	B	G	B	G	←
→	G	B	G	B	
→	G	G	G	G	←

This results in two cases where one child is a girl, out of four remaining cases, for a probability of $\frac{1}{2}$.

3. Given the existing data, we can calculate the mean of the GPA of honors students to be $\mu_H = \frac{4.0+3.7+2.5}{3} = 3.4$, and a standard deviation of

$$\sigma_H = \sqrt{\frac{(4.0 - \mu_H)^2 + (3.7 - \mu_H)^2 + (2.5 - \mu_H)^2}{3}} = 0.648$$

We then find the mean and standard deviation of GPA of non-honor students.

$$\mu_N = \frac{3.8 + 3.3 + 3.0 + 3.0 + 2.7 + 2.2}{6} = 3$$

$$\sigma_N = \sqrt{\frac{(3.8 - \mu_N)^2 + (3.3 - \mu_N)^2 + (3.0 - \mu_N)^2 + (3.0 - \mu_N)^2 + (2.7 - \mu_N)^2 + (2.2 - \mu_N)^2}{6}} = 0.493$$

Last, we calculate the general mean and standard deviation, whether honor students or not:

$$\mu = \frac{4 + 3.7 + 2.5 + 3.8 + 3.3 + 3 + 3 + 2.7 + 2.2}{9} = 3.133$$

$$\sigma^2 = \frac{(4-\mu)^2 + (3.7-\mu)^2 + (2.5-\mu)^2 + (3.8-\mu)^2 + (3.3-\mu)^2 + (3-\mu)^2 + (3-\mu)^2 + (2.7-\mu)^2 + (2.2-\mu)^2}{9}$$

$$\sigma = 0.582$$

We then find the gaussian functions of $P(GPA = x)$, $P(GPA = x \mid H)$ and $P(GPA = x \mid N)$:

$$P(GPA = x) = \frac{1}{\sqrt{2\pi}(0.5)} e^{-\frac{(x-3.1)^2}{0.677}}$$

$$P(GPA = x \mid H) = \frac{1}{\sqrt{2\pi}(0.6)} e^{-\frac{(x-3.4)^2}{0.72}}$$

$$P(GPA = x \mid N) = \frac{1}{\sqrt{2\pi}(0.493)} e^{-\frac{(x-3)^2}{0.486}}$$

Now, find $P(AP \mid H)$, $P(notAP \mid H)$, $P(AP \mid N)$, and $P(notAP \mid N)$ by maximum likelihood estimation for Bernoulli distribution:

$$P(X_{AP}|H) = \prod_{i=1}^{n} {P_H}^{x^{(i)}} (1-P_H)^{1-x^{(i)}}$$

$$P(X_{AP}|N) = \prod_{i=1}^{n} {P_N}^{x^{(i)}} (1-P_N)^{1-x^{(i)}}$$

$$P(H|AP) = \frac{P(AP|H)P(H)}{P(AP)} = \frac{\frac{2}{3} * \frac{1}{3}}{\frac{4}{9}} = \frac{2}{9} * \frac{9}{4} = \frac{1}{2}$$

$$P(H|notAP) = \frac{P(notAP|H)P(H)}{P(AP)} = \frac{\frac{1}{3} * \frac{1}{3}}{\frac{5}{9}} = \frac{1}{9} * \frac{9}{5} = \frac{1}{5}$$

$$P(H|GPA = x) = \frac{P(GPA = x|H)P(H)}{P(GPA = x)}$$

$$= \frac{\frac{1}{\sqrt{2\pi}0.6} e^{-\frac{(x-3.4)^2}{0.72}} * \frac{1}{3}}{\frac{1}{\sqrt{2\pi}0.5} e^{-\frac{(x-3.1)^2}{0.677}}} = (0.277)e^{\frac{(0.043)x^2 + (0.139)x - 0.907}{0.487}} = (0.277)e^{(0.088)x^2 + (0.285)x - 1.86}$$

Now, we must find values where $P(H|AP, GPA = x) = 0.5$ and $P(H|notAP, GPA = x) = 0.5$, where

$$P(H|AP, GPA = x) = P(H|AP) * P(H|GPA = x)$$

$$= \frac{1}{2} P(H|GPA = x) = (0.139)e^{(0.088)x^2 + (0.285)x - 1.86}$$

$$P(H|notAP, GPA = x) = P(H|notAP) * P(H|GPA = x)$$

$$= \frac{1}{5} P(H|GPA = x) = (0.055)e^{(0.088)x^2 + (0.285)x - 1.86}$$

find $value_1$ and $value_2$ where for $value_1 < GPA < value2 : P(H|AP, GPA) \geq 0.5$

$$P(H|AP, GPA = x) = (0.139)e^{(0.088)x^2 + (0.285)x - 1.86} \geq 0.5$$

$$(0.088)x^2 + (0.285)x \geq ln(0.5/0.139) + 1.86$$

$$x(0.285 + (0.088)x) \geq 3.140$$

find $value_1$ and $value_2$ where for $value_1 < GPA < value2 : P(H|notAP, GPA) \geq 0.5$

$$P(H|notAP, GPA = x) = (0.055)e^{(0.088)x^2 + (0.285)x - 1.86} \geq 0.5$$

$$(0.088)x^2 + (0.285)x \geq ln(0.5/0.055) + 1.86$$
$$x(0.285 + (0.088)x) \geq 4.067$$

Our final prediction is:
If AP courses are taken, predict H if the GPA is between 3.140 and 4, and if AP courses are not taken, predict H if the GPA is between 1.987 and 4

4.

$$E[V]E[W] = E[VW] \quad (to\ prove)$$
$$E[V] = \sum_{i=1}^{n} v_i P(v_i), E[W] = \sum_{j=1}^{n} w_i P(w_i) \quad (Definition)$$
$$E[VW] = \sum_{i=1}^{n} v_i w_i P(v_i, w_i) = \sum_{i=1}^{n} v_i P(v_i) w_i P(v_i) \quad (independence)$$
$$= \sum_{i=1}^{n} v_i P(v_i) \sum_{i=1}^{n} w_i P(w_i) \quad (Associativity)$$
$$= E[V]E[W] \ \blacksquare$$