CIS 419 HW3

Devin Stein

November 3, 2016

1 Probability Decision Boundary

1.1

Show that the decision $\hat{y}$ that minimizes the expected loss is equivalent to setting a probability threshold θ and predicted $\hat{y}$ =0 if $p_1 < \theta$ and $\hat{y} = 1$ if $p_1 \geq \theta$.

We know the expected loss of decision $\hat{y} = 0$, $\hat{y} = 1$, and the relationship between the probabilities of these two functions:

$$E(\hat{y} = 0) = p_0 * c_1 + p_1 * c_2$$

$$E(\hat{y} = 1) = p_0 * c_3 + p_1 * c_4$$

$$p_0 = 1 - p_1$$

As the costs associated with each decision is fixed, we have three equations for two unknowns p_0 and p_1.The point at which these equations intersects will minimize the overall expected loss. Substituting gives us:

$$E(\hat{y} = 0) = (1 - p_1) * c_1 + p_1 * c_2$$

$$E(\hat{y} = 1) = (1 - p_1 * c_3 + p_1 * c_4$$

By setting these two equations we can find the probability threshold:

$$E(\hat{y} = 0) = E(\hat{y} = 1)$$

$$(1 - p_1) * c_1 + p_1 * c_2 = (1 - p_1) * c_3 + p_1 * c_4$$

$$p_1 = \frac{c_3 - c_1}{c_2 + c_3 - c4} = \theta$$

At $\theta = p_1 = p_2$ we are indifferent to predicting $\hat{y} = 0$ or $\hat{y} = 1$. However, if $\theta > p_1$ we predict $\hat{y} = 0$ else if $\theta \leq p_1$ we predict $\hat{y} = 1$.

1.2

What is the threshold for this loss matrix?

We can find the threshold for this loss matrix by simply substituting the cost values of the matrix into the equation found above:

$$\theta = \frac{c_3 - c_1}{c_2 + c_3 - c4}$$

$$\theta = \frac{5 - 0}{10 + 5 - 0}$$

$$\theta = \frac{1}{3}$$

2 Double Counting the Evidence

The Naive Bayes Decision Rule is: $argmax_{y_k} p(Y = y_k) * p(X_1 = x_1|Y = y_k) * p(X_2 = x_2|Y = y_k)$

2.1

What is the expected error rate of naive Bayes if it uses only attribute X_1? What if it uses only X_2?

The expected error rate is $p(X_1, X_2, Y | Y \neq \hat{Y}(X_1, X_2))$. For X_1, the expected error rate will be:

$$p(Y = T) * p(X_1 = F | Y = T) + p(Y = F) * p(X_1 = T | Y = F)$$

From the given probabilities, we can solve for the unknown probabilities.

$$p(Y = F) = 1 - P(Y = T) = 0.5$$

$$p(X_1 = F | Y = T) = 1 - p(X_1 = T | Y = T) = .2$$

$$p(X_1 = T | Y = F) = 1 - p(X_1 = F | Y = F) = .3$$

Substituting these values in:

$$0.5 * 0.2 + 0.5 * .3 = 0.25$$

Thus, our expected error rate for $X_1 = 0.25$. Using the same analysis for X_2 we get:

$$p(Y = T) * p(X_2 = F | Y = T) + p(Y = F) * p(X_2 = T | Y = F)$$

$$p(X_2 = F | Y = T) = 1 - p(X_2 = T | Y = T) = .1$$

$$p(X_2 = T | Y = F) = 1 - p(X_2 = F | Y = F) = .5$$

$$0.5 * 0.5 + 0.5 * .1 = 0.3$$

Giving an expected error rate of 0.3 for X_2.

2.2

Show that if naive Bayes uses both attributes, X_1 and X_2, the error rate is 0.235. which is better than if using only a single attributes (X_1 or X_2)

In order to calculate the error rate, we need to first calculate the predictions for Naive Bayes using X_1 and X_2.

$$p(X_1 = T, X_2 = T, Y = T) = 0.5 * 0.8 * 0.5 = 0.2$$

$$p(X_1 = T, X_2 = T, Y = F) = 0.5 * 0.3 * 0.1 = 0.015$$

$$p(X_1 = T, X_2 = F, Y = T) = 0.5 * 0.8 * 0.5 = 0.2$$

$$p(X_1 = T, X_2 = F, Y = F) = 0.5 * 0.3 * 0.9 = 0.135$$

$$p(X_1 = F, X_2 = T, Y = T) = 0.5 * 0.2 * 0.5 = 0.05$$

$$p(X_1 = F, X_2 = T, Y = F) = 0.5 * 0.7 * 0.1 = 0.035$$

$$p(X_1 = F, X_2 = F, Y = T) = 0.5 * 0.2 * 0.5 = 0.05$$

$$p(X_1 = F, X_2 = F, Y = F) = 0.5 * 0.7 * 0.9 = 0.315$$

By using the predictions, we can calculate the expected error rate for both attributes:

$$P(X_1 = T, X_2 = T, Y = F) + P(X_1 = T, X_2 = F, Y = F) + P(X_1 = F, X_2 = T, Y = F) + P(X_1 = F, X_2 = F, Y = T)$$

$$0.015 + 0.135 + 0.035 + 0.05 = 0.235$$

This expected error rate is better than both the expected error rates of using X_1 or X_2 alone.

2.3

Now suppose that we create new attribute X_3 that is an exact copy of X_2. What is the expected error of naive Bayes now?

As the only Recalculating our predictions we get:

$$p(X_1 = T, X_2 = T, Y = T) = 0.5 * 0.8 * 0.5 * 0.5 = 0.1$$
$$p(X_1 = T, X_2 = T, Y = F) = 0.5 * 0.3 * 0.1 * 0.1 = 0.0015$$
$$p(X_1 = T, X_2 = F, Y = T) = 0.5 * 0.8 * 0.5 * 0.5 = 0.1$$
$$p(X_1 = T, X_2 = F, Y = F) = 0.5 * 0.3 * 0.9 * 0.9 = 0.1215$$
$$p(X_1 = F, X_2 = T, Y = T) = 0.5 * 0.2 * 0.5 * 0.5 = 0.025$$
$$p(X_1 = F, X_2 = T, Y = F) = 0.5 * 0.7 * 0.1 * 0.1 = 0.0035$$
$$p(X_1 = F, X_2 = F, Y = T) = 0.5 * 0.2 * 0.5 * 0.5 = 0.025$$
$$p(X_1 = F, X_2 = F, Y = F) = 0.5 * 0.7 * 0.9 * 0.9 = 0.2835$$

Plugging these new values into our expected error rate function gives us:

$$0.0015 + 0.1 + 0.0035 + .025 = 0.13$$

2.4

Briefly explain what is happening with naive Bayes

By using Naive Bayes, we are assuming all attributes are conditionally independent from Y. However, X_2 and X_3 are not independent, they are equivalent. Thus, the information added is false and the algorithm thinks it is getting a lower error rate.

2.5

Does logistic regression suffer from the same problem? Briefly explain why...

No it does not. Logistic regression does not assume conditional independence.

3 Training the Best Classifier

Given that we were trying to classify labeled data and had a training set of between 5,000-6,000 points, the best classifiers were likely going to be a K-Nearest Neighbor, Ensemble like AdaBoost, SVC, or LinearSVC. After training and testing all of these classifiers on the challenge training set and testing their prediction power, SVC reigned champion with 99.35% accuracy. I did not need to tune any parameters to achieve these results.