CIS 519: Introduction to Machine Learning
Homework 3

Archith Shivanagere Muralinath(PennID: 82629708, PennKey: archith)

1 PART 1: Problem Set

Problem 1: Probability Decision Boundary

(a)

We know that, $p_0 + p_1 = 1$
$\implies p_0 = 1 - p_1$

Let l_{ij} represent the elements of given loss matrix. Then, we have

$Cost(\hat{y} = 0) = p_0 l_{00} + p_1 l_{01}$

$Cost(\hat{y} = 1) = p_0 l_{10} + p_1 l_{11}$

Substituting for p_0 in above 2 equations, we get

$Cost(\hat{y} = 0) = (1 - p_1) l_{00} + p_1 l_{01}$

$Cost(\hat{y} = 1) = (1 - p_1) l_{10} + p_1 l_{11}$

Amongst all the data points, there will always be one point, where we will have equal costs for predicting true or false. As we are minimizing the cost, deviating from this point will provide us one or the other prediction.

At some point, where the cost is equal for predicting true or false, we have

$(1 - p_1) l_{00} + p_1 l_{01} = (1 - p_1) l_{10} + p_1 l_{11}$

Solving for p_1 gives us

$p_1 = \frac{l_{10} - l_{00}}{l_{01} + l_{10} - (l_{00} + l_{11})}$

Suppose $p_1 = \theta$, then the prediction will be undistinguished.
Therefore, in case $p_1 < \theta$, we are better off if we predict $\hat{y} = 0$ and in case $p_1 \geq \theta$, we are better off predicting $\hat{y} = 1$.

(b)

We know the costs of predictions are given by,

$Cost(\hat{y} = 0) = p_0 l_{00} + p_1 l_{01}$

$Cost(\hat{y} = 1) = p_0 l_{10} + p_1 l_{11}$

Substituting the loss matrix elements, we get

$Cost(\hat{y} = 0) = 10 p_1$

$Cost(\hat{y} = 1) = 5 p_0$

Substituting for p_0, we get

$10 p_1 = 5(1 - p_1)$

$p_1 = \frac{1}{3}$

Therefore, the threshold for given loss matrix is

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

Problem 2: Double counting the evidence

Naive Bayes decision rule is given by

$$h(x) = argmax_{y_k} logP(Y = y_k) + \sum_{j=1}^{d} logP(X_j = x_j | Y = y_k)$$

We are given

$p(Y = T) = 0.5$
$\implies p(Y = F) = 0.5$

Also,

$p(X_1 = T | Y = T) = 0.8$
$p(X_2 = T | Y = T) = 0.5$
$p(X_1 = F | Y = F) = 0.7$
$p(X_2 = F | Y = F) = 0.9$

Attribute X_1 provides slightly stronger evidence about the class label than X_2, and assume X_1 and X_2 are truly independent given Y.

To substitute values in Bayes rule and solve, we need $p(X = T)$. Therefore, marginalizing will give us:

$p(X = T) = p(X = T|Y = T) \times p(Y = T) + p(X = T|Y = F) \times p(Y = F)$

Substituting values in the above equation for $p(X_1 = T)$,

$p(X_1 = T) = 0.8 \times 0.5 + 0.3 \times 0.5$
$p(X_1 = T) = 0.55$
$\implies p(X_1 = F) = 0.45$

Similarly for $p(X_2 = T)$,

$p(X_2 = T) = 0.5 \times 0.5 + 0.1 \times 0.5$
$p(X_2 = T) = 0.3$
$\implies p(X_2 = F) = 0.7$

Problem (2a)

Bayes rule is given by

$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$

Therefore, $p(Y = T|X = T) = \frac{p(X=T|Y=T) \times p(Y=T)}{p(X=T)}$
From this equation, we can substitute values to get

$p(Y = T|X_1 = T) = \frac{0.8 \times 0.5}{0.55} = 0.7272$

$\implies p(Y = F|X_1 = T) = 0.2727$

$p(Y = T|X_1 = F) = \frac{0.2 \times 0.5}{0.45} = 0.2222$

$\implies p(Y = F|X_1 = F) = 0.7777$

With these values, to minimize the cost, we predict $\hat{y} = X_1$. So, error rate is given by

$expectedErrorRate = p(Y = T|X_1 = F) + p(Y = F|X_1 = T)$
$expectedErrorRate = 0.1 + 0.15 = 0.25$

Repeating this for X_2, we get

$p(Y = T|X_2 = T) = \frac{0.5 \times 0.5}{0.3} = 0.8333$

$\implies p(Y = F|X_2 = T) = 0.1666$

$p(Y = T|X_2 = F) = \frac{0.5 \times 0.5}{0.7} = 0.3571$

$\implies p(Y = F|X_2 = F) = 0.6428$

With these values, to minimize the cost, we predict $\hat{y} = X_2$. So, error rate is given by

$expectedErrorRate = p(Y = T|X_2 = F) + p(Y = F|X_2 = T)$
$expectedErrorRate = 0.25 + 0.05 = 0.3$

Problem (2b)

We are given that, X_1 and X_2 are truly independent given Y. If we use both attributes X_1 and X_2, we will have 4 different combinations of these, which will lead us to 4 predictions for Y. Therefore, we can write

$p(X_1 \wedge X_2|Y) = p(X_1|Y) \times p(X_2|Y) \times p(Y) \times \alpha$

———————————**combination 1:** $X_1 = F \wedge X_2 = F$———————————

$p(Y = T|X_1 = F \wedge X_2 = F) = p(X_1 = F|Y = T) \times p(X_2 = F|Y = T) \times p(Y = T) \times \alpha$
$p(Y = T|X_1 = F \wedge X_2 = F) = 0.2 \times 0.5 \times 0.5 \times \alpha = 0.05$

$p(Y = F|X_1 = F \wedge X_2 = F) = p(X_1 = F|Y = F) \times p(X_2 = F|Y = F) \times p(Y = F) \times \alpha$
$p(Y = F|X_1 = F \wedge X_2 = F) = 0.7 \times 0.9 \times 0.5 \times \alpha = 0.315$

———————————**combination 2:** $X_1 = F \wedge X_2 = T$———————————

$p(Y = T|X_1 = F \wedge X_2 = T) = p(X_1 = F|Y = T) \times p(X_2 = T|Y = T) \times p(Y = T) \times \alpha$
$p(Y = T|X_1 = F \wedge X_2 = T) = 0.2 \times 0.5 \times 0.5 \times \alpha = 0.05$

$p(Y = F|X_1 = F \wedge X_2 = T) = p(X_1 = F|Y = F) \times p(X_2 = T|Y = F) \times p(Y = F) \times \alpha$
$p(Y = F|X_1 = F \wedge X_2 = T) = 0.3 \times 0.9 \times 0.5 \times \alpha = 0.035$

combination 3: $X_1 = T \wedge X_2 = F$

$p(Y = T|X_1 = T \wedge X_2 = F) = p(X_1 = T|Y = T) \times p(X_2 = F|Y = T) \times p(Y = T) \times \alpha$
$p(Y = T|X_1 = T \wedge X_2 = F) = 0.8 \times 0.5 \times 0.5 \times \alpha = 0.2$

$p(Y = F|X_1 = T \wedge X_2 = F) = p(X_1 = T|Y = F) \times p(X_2 = F|Y = F) \times p(Y = F) \times \alpha$
$p(Y = F|X_1 = T \wedge X_2 = F) = 0.3 \times 0.9 \times 0.5 \times \alpha = 0.135$

combination 4: $X_1 = T \wedge X_2 = T$

$p(Y = T|X_1 = T \wedge X_2 = T) = p(X_1 = T|Y = T) \times p(X_2 = T|Y = T) \times p(Y = T) \times \alpha$
$p(Y = T|X_1 = T \wedge X_2 = T) = 0.8 \times 0.5 \times 0.5 \times \alpha = 0.2$

$p(Y = F|X_1 = T \wedge X_2 = T) = p(X_1 = T|Y = F) \times p(X_2 = T|Y = F) \times p(Y = F) \times \alpha$
$p(Y = F|X_1 = T \wedge X_2 = T) = 0.3 \times 0.1 \times 0.5 \times \alpha = 0.015$

Based on the values calculated above, we can say that

If $X_1 = F$ and $X_2 = F$, then $\hat{y} = F$
If $X_1 = F$ and $X_2 = T$, then $\hat{y} = T$
If $X_1 = T$ and $X_2 = F$, then $\hat{y} = T$
If $X_1 = T$ and $X_2 = T$, then $\hat{y} = T$

Therefore, now by substituting the other values in $expectedErrorRate$ equation, we get

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

Problem (2c)

We are given that, X_1 and X_2 are truly independent given Y, and we add a new attribute X_3 that is an exact copy of X_2. If we use both attributes X_1 and X_2, we will have 4 different combinations of these, which will lead us to 4 predictions for Y. Therefore, we can write

$p(X_1 \wedge X_2 \wedge X_3|Y) = p(X_1|Y) \times p(X_2|Y) \times p(X_3|Y) \times p(Y) \times \alpha$

As we know X_3 is an exact copy of X_2, we can rewrite this equation as

$p(X_1 \wedge X_2 \wedge X_3|Y) = p(X_1|Y) \times p^2(X_2|Y) \times p(Y) \times \alpha$

Similar to problem 2b, substituting the values

———————————**combination 1:** $X_1 = F \wedge X_2 = F \wedge X_3 = F$———————————

$p(Y = T|X_1 = F \wedge X_2 = F \wedge X_3 = F) = 0.2 \times 0.5 \times 0.5 \times 0.5 \times \alpha = 0.025$

$p(Y = F|X_1 = F \wedge X_2 = F \wedge X_3 = F) = 0.7 \times 0.9 \times 0.9 \times 0.5 \times \alpha = 0.2835$

———————————**combination 2:** $X_1 = F \wedge X_2 = T \wedge X_3 = T$———————————

$p(Y = T|X_1 = F \wedge X_2 = T \wedge X_3 = T) = 0.2 \times 0.5 \times 0.5 \times 0.5 \times \alpha = 0.025$

$p(Y = F|X_1 = F \wedge X_2 = T \wedge X_3 = T) = 0.7 \times 0.1 \times 0.1 \times 0.5 \times \alpha = 0.0035$

———————————**combination 3:** $X_1 = T \wedge X_2 = F \wedge X_3 = F$———————————

$p(Y = T|X_1 = T \wedge X_2 = F \wedge X_3 = F) = 0.8 \times 0.5 \times 0.5 \times 0.5 \times \alpha = 0.1$

$p(Y = F|X_1 = T \wedge X_2 = F \wedge X_3 = F) = 0.3 \times 0.9 \times 0.9 \times 0.5 \times \alpha = 0.1215$

———————————**combination 4:** $X_1 = T \wedge X_2 = T \wedge X_3 = T$———————————

$p(Y = T|X_1 = T \wedge X_2 = T \wedge X_3 = T) = 0.8 \times 0.5 \times 0.5 \times 0.5 \times \alpha = 0.1$

$p(Y = F|X_1 = T \wedge X_2 = T \wedge X_3 = T) = 0.3 \times 0.1 \times 0.1 \times 0.5 \times \alpha = 0.0015$

Based on the values calculated above, we can say that

If $X_1 = F$, $X_2 = F$ and $X_3 = F$, then $\hat{y} = F$
If $X_1 = F$, $X_2 = T$ and $X_3 = T$, then $\hat{y} = T$
If $X_1 = T$, $X_2 = F$ and $X_3 = F$, then $\hat{y} = F$
If $X_1 = T$, $X_2 = T$ and $X_3 = T$, then $\hat{y} = T$

Therefore, now by substituting the other values in *expectedErrorRate* equation, we get

$expectedErrorRate = 0.025 + 0.0035 + 0.1 + 0.0015$
$expectedErrorRate = 0.13$

Problem (2d)

After adding X_3, which is a copy of X_2, they are not independent as they have the same values i.e, they are the same feature. It seems that we are adding data but we are not adding any real data, we are just copying an existing feature to a new one.

Problem (2e)

Logistic Regression does not suffer from the same problem. In this case, the weights will be distributed equally and since we are not actually adding any new data, the accuracy will remain the same.

Problem 3: Reject Option

Problem (3a)

We are given that $p(y = 1|x) = 0.2$.
This exercise is similar to exercise 1. Therefore, incorporating the deductions from it, we have

$p_0 + p_1 = 1$
$Cost(\hat{y} = 0) = 0.2 \times 10 = 2$
$Cost(\hat{y} = 1) = 0.8 \times 10 = 8$

Therefore, predicting 0 minimizes the expected loss for that instance.

Problem (3b)

We are given that $p(y = 1|x) = 0.4$.
Similar to previous problem,

$Cost(\hat{y} = 0) = 0.4 \times 10 = 4$
$Cost(\hat{y} = 1) = 0.6 \times 10 = 6$

Between these two, predicting 0 minimizes the expected loss for that instance. But, in this scenario, predicting either 0 or 1 have an expected loss of greater than 3. We are given a reject option, which has less cost compared to these two. Therefore, REJECTing the instance will minimize the expected loss.

Problem (3c)

This exercise too is similar to problem 1. Therefore, incorporating the deductions from it, we have

$Cost(\hat{y} = 0) = p_0 l_{00} + p_1 l_{01}$

$Cost(\hat{y} = 1) = p_0 l_{10} + p_1 l_{11}$

But in this problem, we are given a reject option. Modifying the above equations, we get

$Cost(\hat{y} = 0) = min(p_0 l_{00} + p_1 l_{01},\, l_{reject})$

$Cost(\hat{y} = 1) = min(p_0 l_{10} + p_1 l_{11},\, l_{reject})$

Further deductions will give us,

$p_0 l_{00} + p_1 l_{01} = l_{reject}$
$p_0 l_{10} + p_1 l_{11} = l_{reject}$

$$\implies l_{00} + p_1 \times (l_{01} - l_{00}) = l_{reject}$$
$$l_{10} + p_1 \times (l_{11} - l_{10}) = l_{reject}$$

$\theta_0 = \frac{l_{reject} - l_{00}}{l_{01} - l_{00}}$

$\theta_1 = \frac{l_{reject} - l_{10}}{l_{11} - l_{10}}$

Problem (3d)

From previous problem, we have

$Cost(\hat{y} = 0) = min(10p_1, 3)$
$Cost(\hat{y} = 1) = min(5p_0, 3)$

We get,

$10\theta_0 = 3$
$$\implies \theta_0 = 0.3$$

$(1 - \theta_1)5 = 3$
$$\implies \theta_1 = 0.4$$