Name: Yihan Zhou
GT account: yzhou376@gatech.edu
GT number: 903053761

CS 4641 Machine Learning
HW3

Problems Set

1.(a). We do not know the exact result of y. Suppose the probability of predicting $\hat{y} = 1$ is p. Then the overall loss of predicting:
$\hat{y} = 0$:

$$0 \cdot (1 - p) + 10 \cdot p = 10 \cdot p$$

$\hat{y} = 1$:

$$0 \cdot p + 5 \cdot (1 - p) = 5 \cdot (1 - p)$$

When predicting $\hat{y} = 0$, we need the loss of predicting 0 smaller than predicting 1, and vice versa. By the inequality, we get a threshold of p. The probability that $\hat{y} = 1$ is p_1. Thus we have a threshold of p_1. As p_1 increases, cost of predicting 0 increases and cost of predicting 1 decreases, and vice versa. Thus, we set a threshold θ, and predicting $\hat{y} = 0$ if $p_1 < \theta$ and $\hat{y} = 1$ if $p_1 \geq \theta$.

1.(b). When predicting $\hat{y} = 0$, we need cost of 0 smaller than cost of 1.

$$10 \cdot p \leq 5 \cdot (1 - p)$$

$$10 \cdot p \leq 5$$

$$p \leq \frac{1}{3}$$

Thus threshold is $\frac{1}{3}$.

2.(a). $errorrate(X_1) = p(X_1 = F \bigwedge Y = T) + p(X_1 = T \bigwedge Y = F) = p(Y = T)p(X_1 = F \mid Y = T) + p(Y = F)p(X_1 = T \mid Y = F) = 0.5 \cdot (1 - 0.8) + 0.5 \cdot (1 - 0.7) = 0.25$
$errorrate(X_2) = p(X_2 = F \bigwedge Y = T) + p(X_2 = T \bigwedge Y = F) = p(Y = T)p(X_2 = F \mid Y = T) + p(Y = F)p(X_2 = T \mid Y = F) = 0.5 \cdot (1 - 0.5) + 0.5 \cdot (1 - 0.9) = 0.3$

2.(b). $p(X_1 = T, X_2 = F, Y = F) = p(Y = F)p(X_1 = T \mid Y = F)p(X_2 = F \mid Y = F) = 0.5 \cdot 0.3 \cdot 0.9 = 0.135$
$p(X_1 = T, X_2 = F, Y = T) = p(Y = T)p(X_1 = T \mid Y = T)p(X_2 = F \mid Y = T) = 0.5 \cdot 0.8 \cdot 0.5 = 0.2$
Thus when we meet the situation that $X_1 = T$ and $X_2 = F$, we predict $Y = T$ since $0.2 > 0.135$.

$p(X_1 = F, X_2 = T, Y = F) = p(Y = F)p(X_1 = F \mid Y = F)p(X_2 = T \mid Y = F) = 0.5 \cdot 0.7 \cdot 0.1 = 0.035$
$p(X_1 = F, X_2 = T, Y = T) = p(Y = F)p(X_1 = F \mid Y = T)p(X_2 = T \mid Y = T) = 0.5 \cdot 0.2 \cdot 0.5 = 0.05$
Thus when we meet the situation that $X_1 = F$ and $X_2 = T$, we predict $Y = T$ since $0.05 > 0.035$.

We will predict incorrectly in the following 4 situations:
$X_1 = F, X_2 = F, \hat{Y} = F, Y = T$:
$p(X_1 = F, X_2 = F, Y = T) = p(Y = T)p(X_1 = F \mid Y = T)p(X_2 = F \mid Y = T) = 0.5 \cdot 0.2 \cdot 0.2 = 0.05$
$X_1 = T, X_2 = T, \hat{Y} = T, Y = F$:
$p(X_1 = T, X_2 = T, Y = F) = p(Y = F)p(X_1 = T \mid Y = F)p(X_2 = T \mid Y = F) = 0.5 \cdot 0.3 \cdot 0.1 = 0.015$
$X_1 = T, X_2 = F, \hat{Y} = T, Y = F$:
0.135 as indicated above.
$X_1 = F, X_2 = T, \hat{Y} = T, Y = F$:
0.035 as indicated above.

In all, the error rate is $0.05 + 0.015 + 0.135 + 0.035 = 0.235$.

2.(c). Since X_3 is the exact same copy of X_2, the situation is equivalent to result dominated by X_2. Thus the new error rate should be the same as the error rate of X_2, which equals 0.3. In computation, within the case that $X_1 = T$, $X_2 = F$, and $X_3 = F$, we predict $Y = F$ and get $errorrate = 0.05 + 0.015 + 0.2 + 0.035 = 0.3$

2.(d). Naive Bayes performs worse due to the introduction of X_3 which breaks the conditional independence assumption. As a consequence the classifier over counts X_2 partially ignoring X_1.

2.(e). Logistic regression does not suffer from the similar problem since the data will only map to some subspace and then by reducing dimension and modifying coefficients we can obtain result as training on independent data set.