# CS4641 HW3

Kung-hsiang, Huang

2016-10-18

# 1 Probability Decision Boundary

(a)
Let $R(a_0|x)$ and $R(a_1|x)$ denotes the risk of predicting $y = 0$ and $y = 1$, respectively.
Since

$$R(a|x) = \sum_y p(y|x)L(y, \hat{y})$$

We would predict $y = 0$ if $R(a_1|x) > R(a_0|x)$

$$\begin{aligned} R(a_1|x) &> R(a_0|x) \\ p(y=1|x)L(1,1) + p(y=0|x)L(0,1) &> p(y=1|x)L(1,0) + p(y=0|x)L(0,0) \\ p_1 \times 0 + p_0 \times 5 &> p_1 \times 10 + p_0 \times 0 \\ (1 - p_1) \times 5 &> p_1 \times 10 \\ p_1 &< \frac{1}{3} = \theta \end{aligned}$$

(b)
Based on (a), we know that the threshold $\theta = \frac{1}{3}$.

# 2 Double Counting The Evidence

(a)
The expected error rate using solely $X_1$ is $P(X_1, Y|Y \neq \hat{Y}(X_1))$. By Naive Bayes, we can derive the following table:

| | $\hat{P}(Y = T, X_1)$ | $\hat{P}(Y = F, X_1)$ | $\hat{Y}$ |
|---|---|---|---|
| $X_1 = T$ | 0.4 | 0.15 | $T$ |
| $X_1 = F$ | 0.1 | 0.35 | $F$ |

Therefore, we have the following:

$$\begin{aligned} &p(X_1, Y|Y \neq \hat{Y}(X_1)) \\ &= p(X_1 = T, Y = F) + P(X_1 = F, Y = T) \\ &= 0.25 \end{aligned}$$

As for $X_2$, the expected error rate can derived with the same technique as the above.

$$\begin{aligned} &p(X_2, Y|Y \neq \hat{Y}(X_2)) \\ &= p(X_2 = T, Y = F) + P(X_2 = F, Y = T) \\ &= 0.3 \end{aligned}$$

(b)
Similar to (a), with Naive Bayes we can derive the following table:

| | $\hat{P}(Y=T, X_1, X_2)$ | $\hat{P}(Y=F, X_1, X_2)$ | $\hat{Y}$ |
|---|---|---|---|
| $X_1 = T, X_2 = T$ | 0.2 | 0.015 | $T$ |
| $X_1 = T, X_2 = F$ | 0.2 | 0.135 | $T$ |
| $X_1 = F, X_2 = T$ | 0.05 | 0.035 | $T$ |
| $X_1 = F, X_2 = F$ | 0.05 | 0.315 | $F$ |

The expected error rate $P(X_1, X_2, Y | Y \neq \hat{Y}(X_1, X_2))$.

$$\begin{aligned}
&P(X_1, X_2, Y | Y \neq \hat{Y}(X_1, X_2)) \\
&= P(X_1 = T, X_2 = T, Y = F) + P(X1 = T, X_2 = F, Y = F) \\
&+ P(X_1 = F, X_2 = T, Y = F) + P(X1 = F, X_2 = F, Y = T) \\
&= 0.235
\end{aligned}$$

(c)
Similar to (a), with Naive Bayes we can derive the following table:

| | $\hat{P}(Y=T, X_1, X_2, X_3)$ | $\hat{P}(Y=F, X_1, X_2, X_3)$ | $\hat{Y}$ |
|---|---|---|---|
| $X_1 = T, X_2 = T, X_3 = T$ | 0.1 | 0.0015 | $T$ |
| $X_1 = T, X_2 = T, X_3 = F$ | 0.1 | 0.0135 | $T$ |
| $X_1 = T, X_2 = F, X_3 = T$ | 0.1 | 0.0135 | $T$ |
| $X_1 = F, X_2 = T, X_3 = T$ | 0.025 | 0.0035 | $T$ |
| $X_1 = T, X_2 = F, X_3 = F$ | 0.1 | 0.1215 | $F$ |
| $X_1 = F, X_2 = T, X_3 = F$ | 0.025 | 0.0315 | $F$ |
| $X_1 = F, X_2 = F, X_3 = T$ | 0.025 | 0.0315 | $F$ |
| $X_1 = F, X_2 = F, X_3 = F$ | 0.025 | 0.2835 | $F$ |

The error is generated according to the true distribution since $X_3$ is a duplication. We have

$$\begin{aligned}
&P(X_1, X_2, Y | Y \neq \hat{Y}(X_1, X_2)) \\
=&P(Y = F, X_1 = T, X_2 = X_3 = T) + P(Y = T, X_1 = T, X_2 = X_3 = F) \\
&+ P(Y = F, X_1 = F, X_2 = X_3 = T) + P(Y = T, X_1 = F, X_2 = X_3 = F) \\
=&P(X_1 = T, X_2 = T, Y = F) + P(X_1 = T, X_2 = F, Y = T) \\
&+ P(X_1 = F, X_2 = T, Y = F) + P(X1 = F, X_2 = F, Y = T) \\
=&0.015 + 0.2 + 0.035 + 0.05 \\
=&0.3
\end{aligned}$$

(d)

Naive Bayes makes the assumption that all the variables are conditionally independent, which does not applied to this case because $X_3$ and $X_2$ are dependent to each other.

(e)

No, because logistic regression does not make conditional independence assumption.

## 1.2 Training The Best Classifier

The reason for choosing SVM is that it is very good at classification. I tried all kinds of kernel, such as linear, RBF, Gaussian, and RBF with C=5 turns out to give me the best result.