ICS635 Homework 2: by Lambert Leong

Problem 1

1. hypothesis - Oliver's blood is at the crime scene. Probability Oliver was at the crime and the other person is AB blood type

$$P(O|D) = \frac{P(D|O)P(O)}{P(D)}$$

Probability Oliver was not at the crime and that of 2 random people, 1 person is type O and the other is type AB

$$P(\tilde{O}|D) = \frac{P(D|\tilde{O})P(\tilde{O})}{P(D)}$$

Likelihood Ratio, Oliver was there/Oliver was not there

$$\frac{P(O|D)}{P(\tilde{O}|D)} = \frac{P(D|O)P(O)}{P(D|\tilde{O})P(\tilde{O})}$$
$$\frac{P(O|D)}{P(\tilde{O}|D)} = \frac{.01}{2 * (.6) * (.01)}$$
$$\frac{P(O|D)}{P(\tilde{O}|D)} = .83$$

The likelihood ratio is not sigificant enought to accept hypothesis, which states that Oliver's blood is at the crime scene

Problem 2

1.

$$\begin{aligned}\widehat{\theta}_{MLE} &= \frac{N_1}{N} \\ &= \frac{1}{1} \\ &= 1\end{aligned}$$

2.

$$\begin{aligned}\widehat{\theta}_{MAP} &= \frac{N_1 + a - 1}{(N_0 + N_1) + a + b - 2} \\ &= \frac{1 + 1 - 1}{0 + 1 + 1 + 1 - 2} \\ &= \frac{1}{1}\end{aligned}$$

3.

$$\begin{aligned}\bar{\theta} &= \frac{N_1 + a}{N_0 + N_1 + a + b} \\ &= \frac{1+1}{0+1+1+1} \\ &= \frac{2}{3}\end{aligned}$$

4.

$$\begin{aligned}p(\theta|x) &\propto p(x|\theta)p(\theta) \\ &= \theta^x (1-\theta)^{1-x} \frac{\Gamma(N_1+\alpha+N_0+\beta))}{\Gamma(N_1+\alpha)\Gamma(N_0+\beta)} \theta^{N_1+\alpha-1}(1-\theta)^{N_0+\beta-1} \\ &= \frac{\Gamma(N_1+\alpha+N_0+\beta))}{\Gamma(N_1+\alpha)\Gamma(N_0+\beta)} \int_0^1 \theta^{x+N_1+\alpha-1}(1-\theta)^{N_0+\beta-x} d\theta \\ &= \frac{\Gamma(N_1+\alpha+N_0+\beta))}{\Gamma(N_1+\alpha)\Gamma(N_0+\beta)} \frac{\Gamma(x+N_1+\alpha)\Gamma(N_0+\beta-x+1)}{\Gamma(N_1+\alpha+N_0+\beta+1)} \\ &= \frac{\Gamma(1+1+0+1))}{\Gamma(1+1)\Gamma(0+1)} \frac{\Gamma(1+1+1)\Gamma(0+1-0+1)}{\Gamma(1+1+0+1+1)} \\ &= \frac{\Gamma(3))}{\Gamma(2)\Gamma(1)} \frac{\Gamma(3)\Gamma(2)}{\Gamma(4)}\end{aligned}$$

Or

$$p(\theta|x) \propto Beta(\theta|N_1 + a, N_0 + b)$$

5.

$$\begin{aligned}p(x_2 = 1|x = 1, \theta) &= Beta(\theta|x_2 = 1, x = 1) \\ &= \theta^{x_2 - 1}(1-\theta)^{x-1}\end{aligned}$$

6.

$$\begin{aligned}\widehat{\theta}_{MLE} &= \frac{N_1}{N} \\ &= \frac{5}{5} \\ &= 1\end{aligned}$$

$$\begin{aligned}\widehat{\theta}_{MAP} &= \frac{N_1 + a - 1}{(N_0 + N_1) + a + b - 2} \\ &= \frac{5 + 1 - 1}{0 + 5 + 1 + 1 - 2} \\ &= \frac{5}{5} \\ &= 1\end{aligned}$$

$$\begin{aligned}\bar{\theta} &= \frac{N_1 + a}{N_0 + N_1 + a + b} \\ &= \frac{5 + 1}{0 + 5 + 1 + 1} \\ &= \frac{6}{7}\end{aligned}$$

$$\begin{aligned}p(\theta|x) &\propto p(x|\theta)p(\theta) \\ &= \theta^x(1-\theta)^{1-x}\frac{\Gamma(N_1+\alpha+N_0+\beta))}{\Gamma(N_1+\alpha)\Gamma(N_0+\beta)}\theta^{N_1+\alpha-1}(1-\theta)^{N_0+\beta-1} \\ &= \frac{\Gamma(N_1+\alpha+N_0+\beta))}{\Gamma(N_1+\alpha)\Gamma(N_0+\beta)}\int_0^1 \theta^{x+N_1+\alpha-1}(1-\theta)^{N_0+\beta-x}d\theta \\ &= \frac{\Gamma(N_1+\alpha+N_0+\beta))}{\Gamma(N_1+\alpha)\Gamma(N_0+\beta)}\frac{\Gamma(x+N_1+\alpha)\Gamma(N_0+\beta-x+1)}{\Gamma(N_1+\alpha+N_0+\beta+1)} \\ &= \frac{\Gamma(5+1+0+1))}{\Gamma(5+1)\Gamma(0+1)}\frac{\Gamma(1+5+1)\Gamma(0+1-0+1)}{\Gamma(5+1+0+1+1)} \\ &= \frac{\Gamma(7))}{\Gamma(6)\Gamma(1)}\frac{\Gamma(7)\Gamma(2)}{\Gamma(8)}\end{aligned}$$

7.

Problem 3

1.

$$\theta_{spam} = \frac{spam}{(spam + normal)} = \frac{3}{(3+4)} = \frac{3}{7}$$
$$\theta_{secret|spam} = \frac{secret\ in\ spam}{(total\ spam)} = \frac{2}{(2+1)} = \frac{2}{3}$$
$$\theta_{secret|non-spam} = \frac{secret\ in\ non-spam}{(total\ non-spam)} = \frac{1}{(2+2)} = \frac{1}{4}$$
$$\theta_{sport|non-spam} = \frac{sport\ in\ non-spam}{(total\ non-spam)} = \frac{2}{(2+2)} = \frac{1}{2}$$
$$\theta_{dollar|spam} = \frac{dollar\ in\ spam}{(total\ spam)} = \frac{1}{(1+2)} = \frac{1}{3}$$

Problem 4

1. The tradeoff between bias and variance is something to consider when training complex models. Increasing the data set may introduce more variance which may introduce more error during training.

2. With equal classess of randomly labeled data, the best misclassification rate is 50%.

   Since the data is randomly labeled, when we leave on out, the classifier is more likely to favor the class which has N data points rather than N-1. It will then misclassify the left out data point and leaving one out will misclassify on every iteration.

Problem 5

1.