COMP 5630 Fall 2022 Assignment 2

Will Humphlett (wah0028)

September 10, 2022

1. Independent Events and Bayes Theorem

   (a) Prove

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

$$\frac{P(A\cap B)}{P(B)} = \frac{\frac{P(B\cap A)P(A)}{P(A)}}{\frac{P(B\cap A)P(A)}{P(A)}+\frac{P(B\cap \neg A)P(\neg A)}{P(\neg A)}}$$

$$\frac{P(A)P(B)}{P(B)} = \frac{P(B)P(A)}{P(B)(P(A)+P(\neg A))} \tag{1}$$

$$P(A)P(B) = \frac{P(B)P(A)}{1}$$

$$P(A)P(B) = P(A)P(B)$$

$$1 = 1$$

   (b) Questions of X, Y, & Z

      i. Is X independent of Y?

$$
\begin{aligned}
P(X = 0) &= .1 + .2 + .1 + .175 = .575 \\
P(X = 1) &= .05 + .1 + .1 + .175 = .425 \\
P(Y = 0) &= .1 + .05 + .1 + .1 = .35 \\
P(Y = 1) &= .2 + .1 + .175 + .175 = .65 \\
P(Z = 0) &= .1 + .05 + .2 + .1 = .45 \\
P(Z = 1) &= .1 + .1 + .175 + .175 = .55
\end{aligned}
\tag{2}
$$

If X is independent of Y, then $P(X \cap Y) = P(X)P(Y)$

$$
\begin{aligned}
P(X = 1 \cap Y = 1) &\neq P(X = 1)P(Y = 1) \\
.1 + .175 &\neq .425 * .65 \\
.275 &\neq .27625
\end{aligned}
\tag{3}
$$

No, X and Y are not independent.

ii. Is X conditionally independent of Y given Z?

$$P(X \cap Y|Z) = \frac{P(X \cap Y \cap Z)}{P(Z)} = \frac{.175}{.55}$$

$$P(X|Z) = \frac{P(X \cap Z)}{P(Z)} = \frac{.1+.175}{.55} \tag{4}$$

$$P(Y|Z) = \frac{P(Y \cap Z)}{P(Z)} = \frac{.175+.175}{.55}$$

If X is independent of Y given Z, then $P(X \cap Y|Z) = P(X|Z)P(Y|Z)$

$$\frac{.175}{.55} = \frac{.1+.175}{.55} * \frac{.175+.175}{.55}$$

$$.3\overline{18} = .5 * .\overline{63} \tag{5}$$

$$.3\overline{18} = .3\overline{18}$$

Yes, X is conditionally independent of Y given Z.

iii. Calculate $P(X \ne Y|Z = 0)$

$$P(X \ne Y|Z = 0) = \frac{P(X \ne Y \cap Z=0)}{P(Z=0)}$$

$$= \frac{.05+.2}{.45} \tag{6}$$

$$= \boxed{.\overline{5}}$$

2. Maximum Likelihood Estimations

 (a) Bernoulli log-likelihood

$$L(\theta) = \prod_{i=1}^{n} P(x_i|\theta)$$

$$= \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$$

$$= \prod_{i=1}^{n} \theta^{x_i} * \prod_{i=1}^{n} (1-\theta)^{1-x_i}$$

$$= \theta^{\sum_{i=1}^{n} x_i} * (1-\theta)^{\sum_{i=1}^{n}(1-x_i)}$$

$$= \theta^{\sum_{i=1}^{n} x_i} * (1-\theta)^{n-\sum_{i=1}^{n} x_i} \tag{7}$$

$$\ell(\theta) = \log L(\theta)$$

$$= \log \left(\theta^{\sum_{i=1}^{n} x_i} * (1-\theta)^{n-\sum_{i=1}^{n} x_i} \right)$$

$$= \log \left(\theta^{\sum_{i=1}^{n} x_i} \right) + \log \left((1-\theta)^{n-\sum_{i=1}^{n} x_i} \right)$$

$$= \sum_{i=1}^{n} (x_i) \log \theta + (n - \sum_{i=1}^{n}(x_i)) \log (1-\theta)$$

(b) Apply Bernoulli log-likelihood

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{1}{\theta} \sum_{i=1}^{n} x_i - \frac{1}{1-\theta}(n - \sum_{i=1}^{n} x_i) = 0$$

$$\frac{1}{\theta} \sum_{i=1}^{n} x_i = \frac{1}{1-\theta}(n - \sum_{i=1}^{n}(x_i))$$

$$\frac{n - \sum_{i=1}^{n}(x_i)}{\sum_{i=1}^{n} x_i} = \frac{1-\theta}{\theta} \tag{8}$$

$$\frac{n}{\sum_{i=1}^{n} x_i} - \not{X} = \frac{1}{\theta} - \not{X}$$

$$\theta = \frac{1}{n} \sum_{i=1}^{n} x_i = \boxed{\overline{x_n}}$$

Max likelihood of $X = (0, 1, 0, 1, 1, 0, 0, 1, 1, 1) = \overline{x_n} = \boxed{0.6}$

(c) Binomial log-likelihood

$$L(\theta) = \prod_{i=1}^{m} P(x_i|\theta)$$

$$= \prod_{i=1}^{m} \left(\frac{n!}{k!(n-k)!} \theta^k (1-\theta)^{n-k}\right)$$

$$= \prod_{i=1}^{m} \frac{n!}{k!(n-k)!} * \prod_{i=1}^{m} \theta^k * \prod_{i=1}^{m} (1-\theta)^{n-k}$$

$$= \prod_{i=1}^{m} \left(\frac{n!}{k!(n-k)!}\right)(\theta^{\sum_{i=1}^{m} k_i} (1-\theta)^{nm - \sum_{i=1}^{m} k_i})$$

$$\ell(\theta) = \log L(\theta)$$

$$= \log \left(\prod_{i=1}^{m} \left(\frac{n!}{k!(n-k)!}\right)(\theta^{\sum_{i=1}^{m} k_i}(1-\theta)^{nm - \sum_{i=1}^{m} k_i})\right)$$

$$= \log \left(\prod_{i=1}^{m} \left(\frac{n!}{k!(n-k)!}\right)\right) + \log \left(\theta^{\sum_{i=1}^{m} k_i}\right) + \log \left((1-\theta)^{nm - \sum_{i=1}^{m} k_i}\right)$$

$$= \log \left(\prod_{i=1}^{m} \left(\frac{n!}{k!(n-k)!}\right)\right) + \left(\sum_{i=1}^{m} k_i\right) \log(\theta) + \left(nm - \sum_{i=1}^{m} k_i\right) \log(1-\theta) \tag{9}$$

(d) Apply Binomial log-likelihood

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{1}{\theta} \sum_{i=1}^{m} k_i - \frac{1}{1-\theta}(nm - \sum_{i=1}^{m} k_i) = 0$$

$$\frac{1}{\theta} \sum_{i=1}^{m} k_i = \frac{1}{1-\theta}(nm - \sum_{i=1}^{m} k_i)$$

$$\frac{1-\theta}{\theta} = \frac{nm - \sum_{i=1}^{m} k_i}{\sum_{i=1}^{m} k_i} \tag{10}$$

$$\frac{1}{\theta} - \not{X} = \frac{nm}{\sum_{i=1}^{m} k_i} - \not{X}$$

$$\theta = \frac{1}{nm} \sum_{i=1}^{m} k_i = \boxed{\frac{\overline{x_m}}{n}}$$

Max likelihood of $Y_1 = (0, 1, 0, 1, 1), Y_2 = (0, 0, 1, 1, 1) = \frac{\overline{x_m}}{n} = \boxed{0.6}$

3

(e) Comparison

The binomial distribution is a combination of a series of Bernoulli distributions. Parts 1 & 3 differ by a constant $(\log{(\prod_{i=1}^{m}{(\frac{n!}{k!(n-k)!})})})$ due to this fact. However, parts 2 & 4 are equal as whatever maxes a series of Bernoulli distributions will also max a single Bernoulli distribution.

3. Implementing Naive Bayes (code at: github.com/wumphlett)

```
Training and Testing Multiple Naive Bayes Implementations
Training Multinomial Naive Bayes
Training Acc: 96.93%
Testing Acc: 98.23%
Execution Time: 0.069s

Training Gaussian Naive Bayes
Training Acc: 90.79%
Testing Acc: 81.01%
Execution Time: 0.473s

Training Complement Naive Bayes
Training Acc: 96.91%
Testing Acc: 98.23%
Execution Time: 0.069s

Training Bernoulli Naive Bayes
Training Acc: 91.83%
Testing Acc: 92.19%
Execution Time: 0.633s

Training Categorical Naive Bayes
Training Acc: 63.55%
Testing Acc: 60.19%
Execution Time: 1.121s
```

Multinomail and Complement implementations of Naive Bayes performed the best with marginal differences between them. They also had the shortest training time.