COMP 5630 Fall 2022 Assignment 2

Will Humphlett (wah0028)

September 10, 2022

1. Independent Events and Bayes Theorem

   (a) Prove

$$
\begin{aligned}
P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\neg A)P(\neg A)} \\
\frac{P(A\cap B)}{P(B)} &= \frac{\frac{P(B\cap A)\cancel{P(A)}}{\cancel{P(A)}}}{\frac{P(B\cap A)\cancel{P(A)}}{\cancel{P(A)}}+\frac{P(B\cap\neg A)\cancel{P(\neg A)}}{\cancel{P(\neg A)}}} \\
\frac{P(A)P(B)}{\cancel{P(B)}} &= \frac{P(B)P(A)}{\cancel{P(B)}(P(A)+P(\neg A))} \\
P(A)P(B) &= \frac{P(B)P(A)}{1} \\
P(A)P(B) &= P(A)P(B) \\
1 &= 1
\end{aligned} \tag{1}
$$

   (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?

$$\begin{aligned} P(X \cap Y|Z) &= \tfrac{P(X \cap Y \cap Z)}{P(Z)} = \tfrac{.175}{.55} \\ P(X|Z) &= \tfrac{P(X \cap Z)}{P(Z)} = \tfrac{.1+.175}{.55} \\ P(Y|Z) &= \tfrac{P(Y \cap Z)}{P(Z)} = \tfrac{.175+.175}{.55} \end{aligned} \tag{4}$$

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

$$\begin{aligned} \tfrac{.175}{.55} &= \tfrac{.1+.175}{.55} * \tfrac{.175+.175}{.55} \\ .3\overline{18} &= .5 * .\overline{63} \\ .3\overline{18} &= .3\overline{18} \end{aligned} \tag{5}$$

Yes, X is conditionally independent of Y given Z.

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

$$\begin{aligned} P(X \neq Y|Z = 0) &= \tfrac{P(X \neq Y \cap Z=0)}{P(Z=0)} \\ &= \tfrac{.05+.2}{.45} \\ &= \boxed{.\bar{5}} \end{aligned} \tag{6}$$

2. Maximum Likelihood Estimations

(a) Bernoulli log-likelihood

$$\begin{aligned} L(\theta) &= \textstyle\prod_{i=1}^{n} P(x_i|\theta) \\ &= \textstyle\prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \textstyle\prod_{i=1}^{n} \theta^{x_i} * \prod_{i=1}^{n}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^{n} x_i} * (1-\theta)^{\sum_{i=1}^{n}(1-x_i)} \\ &= \theta^{\sum_{i=1}^{n} x_i} * (1-\theta)^{n-\sum_{i=1}^{n} x_i} \end{aligned} \tag{7}$$

$$\begin{aligned} \ell(\theta) &= \log L(\theta) \\ &= \log(\theta^{\sum_{i=1}^{n} x_i} * (1-\theta)^{n-\sum_{i=1}^{n} x_i}) \\ &= \log(\theta^{\sum_{i=1}^{n} x_i}) + \log((1-\theta)^{n-\sum_{i=1}^{n} x_i}) \\ &= \textstyle\sum_{i=1}^{n}(x_i)\log\theta + (n - \sum_{i=1}^{n}(x_i))\log(1-\theta) \end{aligned}$$

(b) Apply Bernoulli log-likelihood

$$\begin{aligned}
\frac{\partial \ell(\theta)}{\partial \theta} &= \tfrac{1}{\theta}\textstyle\sum_{i=1}^{n} x_i - \tfrac{1}{1-\theta}(n - \sum_{i=1}^{n} x_i) = 0 \\
\tfrac{1}{\theta}\textstyle\sum_{i=1}^{n} x_i &= \tfrac{1}{1-\theta}(n - \textstyle\sum_{i=1}^{n}(x_i)) \\
\frac{n-\sum_{i=1}^{n}(x_i)}{\sum_{i=1}^{n} x_i} &= \frac{1-\theta}{\theta} \\
\frac{n}{\sum_{i=1}^{n} x_i} - \cancel{1} &= \frac{1}{\theta} - \cancel{1} \\
\theta &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n} x_i = \boxed{\overline{x_n}}
\end{aligned} \tag{8}$$

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

(c) Binomial log-likelihood

$$\begin{aligned}
L(\theta) &= \textstyle\prod_{i=1}^{m} P(x_i|\theta) \\
&= \textstyle\prod_{i=1}^{m}(\frac{n!}{k!(n-k)!}\theta^{k}(1-\theta)^{n-k}) \\
&= \textstyle\prod_{i=1}^{m} \frac{n!}{k!(n-k)!} * \prod_{i=1}^{m} \theta^{k} * \prod_{i=1}^{m}(1-\theta)^{n-k} \\
&= \textstyle\prod_{i=1}^{m}(\frac{n!}{k!(n-k)!})(\theta^{\sum_{i=1}^{m} k_i}(1-\theta)^{nm-\sum_{i=1}^{m} k_i}) \\
\\
\ell(\theta) &= \log L(\theta) \\
&= \log(\textstyle\prod_{i=1}^{m}(\frac{n!}{k!(n-k)!})(\theta^{\sum_{i=1}^{m} k_i}(1-\theta)^{nm-\sum_{i=1}^{m} k_i})) \\
&= \log(\textstyle\prod_{i=1}^{m}(\frac{n!}{k!(n-k)!})) + \log(\theta^{\sum_{i=1}^{m} k_i}) + \log((1-\theta)^{nm-\sum_{i=1}^{m} k_i}) \\
&= \log(\textstyle\prod_{i=1}^{m}(\frac{n!}{k!(n-k)!})) + (\sum_{i=1}^{m} k_i)\log(\theta) + (nm - \sum_{i=1}^{m} k_i)\log(1-\theta)
\end{aligned} \tag{9}$$

(d) Apply Binomial log-likelihood

$$\begin{aligned}
\frac{\partial \ell(\theta)}{\partial \theta} &= \tfrac{1}{\theta}\textstyle\sum_{i=1}^{m} k_i - \tfrac{1}{1-\theta}(nm - \sum_{i=1}^{m} k_i) = 0 \\
\tfrac{1}{\theta}\textstyle\sum_{i=1}^{m} k_i &= \tfrac{1}{1-\theta}(nm - \textstyle\sum_{i=1}^{m} k_i) \\
\frac{1-\theta}{\theta} &= \frac{nm-\sum_{i=1}^{m} k_i}{\sum_{i=1}^{m} k_i} \\
\frac{1}{\theta} - \cancel{1} &= \frac{nm}{\sum_{i=1}^{m} k_i} - \cancel{1} \\
\theta &= \tfrac{1}{nm}\textstyle\sum_{i=1}^{m} k_i = \boxed{\frac{\overline{x_m}}{n}}
\end{aligned} \tag{10}$$

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

(e) Comparison

The binomial distribution is a combination of a series of Bernoulli distributions. Parts 1 & 3 differ by a constant ($\log\left(\prod_{i=1}^{m}\left(\frac{n!}{k!(n-k)!}\right)\right)$) 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
```

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