CS446: Machine Learning, Fall 2017, Homework 1

Name: Lanxiao Bai (lbai5)

Worked individually

Problem 1

Solution: By counting by the categories in the dataset D_n, we have prior probabilities

$$P(y = A) = \frac{\sum_{y \in D_n} 1_{[y=A]}}{n} = \frac{3}{7}$$

$$P(y = B) = \frac{\sum_{y \in D_n} 1_{[y=B]}}{n} = \frac{4}{7}$$

In order to estimate the parameters $\lambda_1^A, \lambda_2^A, \lambda_1^B, \lambda_2^B$, we can apply Maximum Likelihood Estimation(MLE).
Since log function is monotone, instead of likelihood function, we can calculate the log likelihood function

$$\begin{aligned}
\mathcal{L}(\lambda_i \mid D_n) &= \log f(D_n \mid \lambda_i) \\
&= \log \prod_{j=1}^{n} f(x_{ij} \mid \lambda) && \text{(Independence Assumption)} \\
&= \sum_{j=1}^{n} \log f(x_{ij} \mid \lambda)
\end{aligned}$$

Thus,

$$\begin{aligned}
\lambda_{MLE} &= \arg\max_{\lambda_i} f(D_n \mid \lambda_i) \\
&= \arg\max_{\lambda_i} \mathcal{L}(\lambda_i \mid D_n) \\
&= \arg\max_{\lambda_i} \sum_{j=1}^{n} \log f(x_{ij} \mid \lambda_i) \qquad (1)
\end{aligned}$$

Now we plug in given probability distribution from Poisson Distribution

$$P(x_i = x \mid y_i = y) = \frac{e^{-\lambda_i^y} (\lambda_i^y)^x}{x!}$$

into (1) and get that

$$
\begin{aligned}
\lambda_i^y &= \arg\max_{\lambda_i^y} \sum_{\{x_{ij}:y_i=A\}} \log f(x_{ij} \mid \lambda_i^A) + \sum_{\{x_{ij}:y_i=B\}} \log f(x_{ij} \mid \lambda_i^B) \\
&= \arg\max_{\lambda_i^y} (-n_A \lambda_i^y + \log \lambda_i^A \sum_{\{x_{ij}:y_i=A\}} x_{ij} - \sum_{\{x_{ij}:y_i=A\}} \log(x_{ij}!)) \\
&+ (-n_B \lambda_i^y + \log \lambda_i^B \sum_{\{x_{ij}:y_i=B\}} x_{ij} - \sum_{\{x_{ij}:y_i=B\}} \log(x_{ij}!)) \\
&= \arg\max_{\lambda_i^y} \left[-n_A \lambda_i^y + \log \lambda_i^y \sum_{\{x_{ij}:y_i=y\}} x_{ij} - \sum_{\{x_{ij}:y_i=y\}} \log(x_{ij}!) \right]
\end{aligned}
$$

So by solving

$$
\frac{\partial \mathcal{L}(\lambda \mid D_n)}{\partial \lambda} = 0 \Rightarrow -n + \frac{1}{\lambda} \sum_{\{x_{ij}:y_i=y\}} x_{ij} = 0
$$

we get that

$$
\lambda_i^y = \frac{1}{n} \sum_{\{x_{ij}:y_i=y\}} x_{ij} \tag{2}
$$

By checking the 2nd order derivative, we can confirm that at this point $\mathcal{L}$ reaches maximum. Finally, by applying (2) to dataset D_n, we have that

$$
\lambda_1^A = \frac{1}{n} \sum_{\{x_{1j}:y_i=A\}} x_{1j} = 3
$$

$$
\lambda_2^A = \frac{1}{n} \sum_{\{x_{2j}:y_i=A\}} x_{2j} = 6
$$

$$
\lambda_1^B = \frac{1}{n} \sum_{\{x_{1j}:y_i=B\}} x_{1j} = 5
$$

$$
\lambda_2^B = \frac{1}{n} \sum_{\{x_{2j}:y_i=B\}} x_{2j} = 4
$$

Problem 2

Solution:

$$
\begin{aligned}
\frac{\Pr(x_1 = 2, x_2 = 3 \mid y = A)}{\Pr(x_1 = 2, x_2 = 3 \mid y = B)} &= \frac{\Pr(x_1 = 2 \mid y = A)\Pr(x_2 = 3 \mid y = A)}{\Pr(x_1 = 2 \mid y = B)\Pr(x_2 = 3 \mid y = B)} \\
&= \frac{\frac{e^{-\lambda_1^A}(\lambda_1^A)^{x_1}}{x_1!} \frac{e^{-\lambda_2^A}(\lambda_2^A)^{x_2}}{x_2!}}{\frac{e^{-\lambda_1^B}(\lambda_1^B)^{x_1}}{x_2!} \frac{e^{-\lambda_2^B}(\lambda_2^B)^{x_2}}{x_2!}} \\
&= \frac{e^{-\lambda_1^A}(\lambda_1^A)^{x_1} e^{-\lambda_2^A}(\lambda_2^A)^{x_2}}{e^{-\lambda_1^B}(\lambda_1^B)^{x_1} e^{-\lambda_2^B}(\lambda_2^B)^{x_2}}
\end{aligned}
$$

$$
\begin{aligned}
&= e^{\lambda_1^B+\lambda_2^B-\lambda_1^A-\lambda_2^A}\frac{(\lambda_1^A)^{x_1}(\lambda_2^A)^{x_2}}{(\lambda_1^B)^{x_1}(\lambda_2^B)^{x_2}} \\
&= \frac{(\lambda_1^A)^{x_1}(\lambda_2^A)^{x_2}}{(\lambda_1^B)^{x_1}(\lambda_2^B)^{x_2}} \\
&= \frac{3^2\cdot 6^3}{5^2\cdot 4^3} \\
&= 1.215
\end{aligned}
$$

Problem 3

Solution: Naive Bayes makes classifier prediction base on the maximum posterior condition probability in all target classes $\mathcal{Y}$, so that class

$$
\begin{aligned}
K &= \arg\max_{K_i\in\mathcal{Y}} P(y = K_i \mid x_1, x_2) \\
&= \arg\max_{K_i\in\mathcal{Y}} P(x_1, x_2 \mid y = K_i)P(y = K_i) \\
&= \arg\max_{K_i\in\mathcal{Y}} P(x_1 \mid y = K_i)P(x_2 \mid y = K_i)P(y = K_i)
\end{aligned}
$$

Considering that our $\mathcal{Y} = \{A, B\}$, and map A to 1, B to -1, then

$$
\begin{aligned}
K &= \text{sign}(\frac{P(x_1 \mid y = A)P(x_2 \mid y = A)P(y = A)}{P(x_1 \mid y = B)P(x_2 \mid y = B)P(y = B)} - 1) \\
&= \text{sign}(\frac{(\lambda_1^A)^{x_1}(\lambda_2^A)^{x_2}}{(\lambda_1^B)^{x_1}(\lambda_2^B)^{x_2}}\frac{P(y = A)}{P(y = B)} - 1) \\
&= \text{sign}(\frac{3\cdot 3^{x_1}6^{x_2}}{4\cdot 5^{x_1}4^{x_2}} - 1) \qquad (3)
\end{aligned}
$$

Problem 4

Solution: By plugging $x_1 = 2, x_2 = 3$ into equation (3) we get that

$$K = \text{sign}(-0.08875) = 0$$

which means we predict that the post should be class $y = B$, a teacher post.

References

(2017). Poisson distribution.
URL https://en.wikipedia.org/wiki/Poisson_distribution#Maximum_likelihood