CIS 419/519: Homework 4

{Yiwen Tang}

Although the solutions are my own, I consulted with the following people while working on this homework: {Names here}

Document Classification

Representation	k	Test Accuracy
BBoW	0.1	0.8600
CBoW	0.1	0.8608
TF-IDF	1.0	0.8300

What impact does k have? That is, as the value of k goes to infinity, what will happen to $P(y \mid d)$?

k is the hyper-parameter for Laplace smoothing. Its main goal is to prevent any probability to get to 0. As k gets larger, the components in the original $P(v|y)$ (i.e without the smoothing hyperparameter) becomes less and less important in the formula. When k approach infinity, all $P(v|y)$ converge to the value $\frac{1}{|V|}$. As a result, the values for $P(y|d)$ will depend on the value of P(y). This means that the label which appears more frequently will be the predicted label of all documents, in this case label 0.

Threshold table:

# Starting Instances	Starting Valid Acc.	Final Valid Acc.	Final Test Acc.
50	0.5672	0.6252	0.6484
500	0.7528	0.7608	0.7544
5000	0.8200	0.7844	0.7832

Top-K table:

# Starting Instances	Starting Valid Acc.	Final Valid Acc.	Final Test Acc.
50	0.6828	0.6960	0.7052
500	0.7740	0.7144	0.7112
5000	0.8284	0.7512	0.7492

Theory

Multivariate Exponential Naive Bayes

1. Table:

$P(Y = A) = \frac{3}{7}$	$P(Y = B) = \frac{4}{7}$
$\lambda_{A;1} = \frac{1}{2}$	$\lambda_{B;1} = \frac{1}{4}$
$\lambda_{A;2} = \frac{1}{5}$	$\lambda_{B;1} = \frac{1}{3}$

Reasoning: 1) There are 7 seemed examples, 3 with $Y = 3$ and 4 with $Y = 4$. Therefore, the prior probabilities are:

$$P(Y = A) = \frac{3}{7} \quad P(Y = B) = \frac{4}{7}$$

2) To solve for the lambdas, we first calculate a general MLE of the probability term $P(X_i = x|Y = label)$. Then we plug in the seemed examples to calculate for the specific lambdas.

$$\begin{aligned}
P(X_i = x|Y = label) &= e^{-x\lambda_{label;i}}\lambda_{label;i} \\
L(\lambda_{label;i}|X_{i,1}\dots X_{i,n}) &= e^{-\sum_{j=1}^{n} x_{i,j}\lambda_{label;i}}\lambda_{label;i}^{n} \\
log(L(\lambda_{label;i}|X_{i,1}\dots X_{i,n})) &= -\sum_{j=1}^{n} x_{i,j}\lambda_{label;i} + nlog(\lambda_{label;i}) \\
\frac{\partial log(L)}{\partial \lambda_{label;i}} &= -\sum_{j=1}^{n} x_{i,j} + \frac{n}{\lambda_{label;i}} = 0 \\
\sum_{j=1}^{n} x_{i,j} &= \frac{n}{\lambda_{label;i}} \\
\lambda_{label;i} &= \frac{n}{\sum_{j=1}^{n} x_{i,j}}
\end{aligned}$$

For example, there are three 3 examples with label $Y = A$, and their X_1 are 0,4,2, correspondingly. Thus, we can calculate $\lambda_{A;1} = \frac{3}{0+4+2} = \frac{1}{2}$. Similarly, we have:

$$\begin{aligned}
\lambda_{A;2} &= \frac{3}{15} = \frac{1}{5} \\
\lambda_{B;1} &= \frac{4}{16} = \frac{1}{4} \\
\lambda_{A;2} &= \frac{4}{12} = \frac{1}{3}
\end{aligned}$$

2. Based on the assumption of Naive Bayes, X_1 and X_2 are independent with respect to Y. Then we have

$$P(X_1 = 2, X_2 = 3|Y = A) = P(X_1 = 2|Y = A) \times P(X_1 = 2|Y = A)$$
$$P(X_1 = 2, X_2 = 3|Y = B) = P(X_1 = 2|Y = B) \times P(X_1 = 2|Y = B)$$

Then we can calculate them individually by plugging in values.

$$\begin{aligned}
P(X_1 = 2, X_2 = 3|Y = A) &= \frac{1}{2}e^{-1} \times \frac{1}{5}e^{-\frac{3}{5}} \\
P(X_1 = 2, X_2 = 3|Y = B) &= \frac{1}{4}e^{-\frac{1}{2}} \times \frac{1}{3}e^{-1} \\
\frac{P(X_1 = 2, X_2 = 3|Y = A)}{P(X_1 = 2, X_2 = 3|Y = B)} &= \frac{\frac{1}{2}e^{-1} \times \frac{1}{5}e^{-\frac{3}{5}}}{\frac{1}{4}e^{-\frac{1}{2}} \times \frac{1}{3}e^{-1}}
\end{aligned}$$

3. The decision rule can be written as follows:

$$\begin{aligned}
R &= \frac{P(X_1 = 2, X_2 = 3|Y = A)}{P(X_1 = 2, X_2 = 3|Y = B)} \times \frac{P(Y = A)}{P(Y = B)} \\
&= \frac{P(X_1 = 2, X_2 = 3|Y = A)\ P(Y = A)}{P(X_1 = 2, X_2 = 3|Y = B)\ P(Y = B)}
\end{aligned}$$

If $R > 1$, the classifier will predict $Y = A$, else it will predict $Y = B$.

4. For $X_1 = 2$, $X_2 = 3$, the classifier will predict $Y = B$.

$$\begin{aligned}
R &= \frac{P(X_1 = 2, X_2 = 3|Y = A)}{P(X_1 = 2, X_2 = 3|Y = B)} \times \frac{P(Y = A)}{P(Y = B)} \\
&= \frac{0.02019}{0.01859} \times \frac{3}{4} \\
&= 0.814 \leq 1
\end{aligned}$$

Coin Toss

The most likely value of p is 0.632. If a T is recorded in the sequence, there are 2 possibilities: 1) We get T at the first time, 2) or we get H first, and the re-toss gives our T. These two events are independent. If a H is recorded in the sequence, there is only 1 possibility, which is two tosses of H happening together. Given $P(H) = p$ and $P(T) = 1 - p$, we have

$$\begin{aligned}
L &= P(T \cup HT)^6 \ P(HH)^4 \\
&= (P(T) + P(HT))^6 \ P(HH)^4 \\
&= (1 - p^2)^6 p^8 \\
log(L) &= 6log(1 - p^2) \ 8log(p) \\
\frac{\partial log(L)}{\partial p} &= \frac{6}{1 - p^2}(-2p) + \frac{8}{p} = 0 \\
p^2 &= \frac{8}{20} \\
p &= \pm 0.632
\end{aligned}$$

Since $p \in [0, 1]$, $p = 0.632$.