1 A simple but relevant example

Suppose we have two input variables X_1, X_2, and one output variable Y. We further assume that

- $X_1 \in \{0, 1\}$
- $X_1 \in \{0, 1, 2\}$
- $Y \in \{0, 1\}$
- $P(x_1, x_2, y) = P(y)P(x_1|y)P(x_2|y)$ (Naive Bayes Assumption)

Here are a few samples for the data set:

X_1	X_2	Y
1	1	0
1	0	1
0	2	1
1	1	0
	$\vdots$	

1.1 How many parameters are needed to estimate the joint likelihood?

- $P(Y = 0)$ and $P(Y = 1)$
- $P(X_1 = 0|Y = 0)$ and $P(X_1 = 1|Y = 0)$
- $P(X_1 = 0|Y = 1)$ and $P(X_1 = 1|Y = 1)$
- $P(X_2 = 0|Y = 0)$, $P(X_2 = 1|Y = 0)$ and $P(X_2 = 2|Y = 0)$
- $P(X_2 = 0|Y = 1)$, $P(X_2 = 1|Y = 1)$ and $P(X_2 = 2|Y = 1)$

So, all in all, we need 12 parameters.

1.2 What are the M.L.E. for the above parameters?

Directly from the notes we have:

$$P(Y = c) = \frac{\#times[Y = c]}{n}$$

$$P(X_i = j|Y = k) = \frac{\#times[Y = k, X_i = j]}{\#times[Y = k]}$$

In Matlab, the code would look something like:

```
sum(Y==0)/n %Computing P(Y=0)
sum( (X1==0).*(Y==1)) / sum(Y==1)  %Computing P(X1=0|Y=1)
```

1.3 What are the M.A.P. for the above parameters, assuming a Dirichlet prior with hyperparameters $\alpha = 2$ on each of the distributions?

We do not go through the derivation here, but we get:

$$P(Y = c) = \frac{\#times[Y = c] + 1}{n + \#classes[Y]}$$

$$P(X_i = j|Y = k) = \frac{\#times[Y = k, X_i = j] + 1}{\#times[Y = k] + \#classes[X_i]}$$

For our simple data set we get:

$$P(Y = 1) = \frac{\#times[Y = 1] + 1}{n + 2}$$

$$P(X_2 = 0|Y = 1) = \frac{\#times[Y = 1, X_2 = 0] + 1}{\#times[Y = 1] + 3}$$

1.4 What is the predicted class Y^* for input: $(X_1, X_2) = (1, 2)$?

$$\begin{aligned}
Y^* &= \arg\max_c \; P(Y = c|X_1 = 1, X_2 = 2) \\
&= \arg\max_c \; \frac{P(Y = c, X_1 = 1, X_2 = 2)}{P(X_1 = 1, X_2 = 2)} \\
&= \arg\max_c \; P(Y = c, X_1 = 1, X_2 = 2) \qquad (P(X_1 = 1, X_2 = 2) \;\text{ is constant}) \\
&= \arg\max_c \; P(Y = c)P(X_1 = 1|Y = c)P(X_2 = 2|Y = c) \qquad \text{(Naive Bayes)}
\end{aligned}$$

1.5 What is the likelihood of the following data set?

X_1	X_2	Y
1	0	0
1	1	0
0	1	1
1	2	1

$$L = P_1 P_2 P_3 P_4$$

Where we have

$$\begin{aligned}
P_1 &= P(X_1 = 1, X_2 = 0, Y = 0) = P(Y = 0)P(X_1 = 1|Y = 0)P(X_2 = 0|Y = 0) \\
P_2 &= P(X_1 = 1, X_2 = 1, Y = 0) = P(Y = 0)P(X_1 = 1|Y = 0)P(X_2 = 1|Y = 0)
\end{aligned}$$

$$\begin{aligned} P_3 &= P(X_1 = 0, X_2 = 1, Y = 1) = P(Y = 1)P(X_1 = 0|Y = 1)P(X_2 = 1|Y = 1) \\ P_4 &= P(X_1 = 1, X_2 = 2, Y = 1) = P(Y = 1)P(X_1 = 1|Y = 1)P(X_2 = 2|Y = 1) \end{aligned}$$

1.6 Assuming the same priors as in 1.3, what is $P(X_1 = 1, X_2 = 2, Y = 0|D)$?

Following the derivation in the notes and in section 4.5.3 (p.131) we get new estimates:

$$P(Y = c) = \frac{\#times[Y = c] + 2}{n + 2 \cdot \#classes[Y]}$$

$$P(X_i = j|Y = k) = \frac{\#times[Y = k, X_i = j] + 2}{\#times[Y = k] + 2 \cdot \#classes[X_i]}$$

The rest of the calculations are exactly the same (for likelihood, and for predictions).

2 Numerical results for the assignment

You should also check against one another to make sure your answers are correct, for example you might add or remove some examples from the training and tests sets and see if your answers remain the same.

Q1

Accuracy: 89.4907%

Q2

Accuracy: 67.6080%

Q3

Training set likelihood: -6.3227e+004

Test set likelihood: -Inf

Q4

- Accuracy: 89.5062%
- Training set likelihood: -6.3238e+004
- Test set likelihood: -6.3379e+004

Q5

Up to you!