4.18

Intro

In this exercise we will perform posterior prediction with mixed features Naive Bayes model. The exercise has nothing special in particular. I suspect that the main reason that it is on the book is to give an example of how flexible Naive Bayes can be. There is one interesting pattern that is particular to this example, which will be the focus of item *d*.

Solution

a)

We need to compute the posterior $p(y|x_1 = 0, x_2 = 0)$. We know the prior $y = Mu(y|\pi, 1) = Cat(y|\pi) = \pi$ and the features conditional distributions. Using Bayes rule

$$\begin{aligned}
p(y|x_1 = 0, x_2 = 0) = \frac{p(x_1 = 0, x_2 = 0|y)p(y)}{p(x_1 = 0, x_2 = 0)} = \\
\frac{p(x_1 = 0, x_2 = 0|y = c)\pi_c}{\sum_{c'} \pi_{c'} p(x_1 = 0, x_2 = 0|y = c')} = \\
\frac{p(x_1 = 0|y = c)p(x_2 = 0|y = c)\pi_c}{\sum_{c'} \pi_{c'} p(x_1 = 0|y = c')p(x_2 = 0|y = c')}
\end{aligned} \tag{1}$$

where in the last step we used the Naive Bayes property. Now, let's calculate the features conditional distributions:

$$\begin{aligned}
&p(x_1 = 0|y = c)p(x_2 = 0|y = c) = Ber(x_1 = 0|\theta_c)\mathcal{N}(x_2 = 0|\mu_c, \sigma_c^2) = \\
&0.5\frac{1}{\sqrt{(2\pi)}} exp\left[-\frac{1}{2}(0 - \mu_c)^2\right] = \\
&0.5\frac{1}{\sqrt{(2\pi)}} exp\left[-\frac{1}{2}(\mu_c)^2\right] = (0.121; 0.199; 0.121)
\end{aligned} \tag{2}$$

With the features conditionals and the priors, we can calculate the joint distribution and the normalization constant:

$$\begin{aligned}
p(x_1 = 0, x_2 = 0, y = c) = (0.0605; 0.04975; 0.03025) \\
p(x_1 = 0, x_2 = 0) = 0.0605 + 0.04975 + 0.03025 = 0.141
\end{aligned} \tag{3}$$

Therefore, the posterior vector is:

$$p(y|x_1 = 0, x_2 = 0) = (0.431, 0.354, 0.215) \tag{4}$$

b)

Know we will calculate the posterior knowing only x_1. This will not be difficult since we know the distribution of $x_1|y$.

$$p(y|x_1 = 0) = \frac{p(x_1 = 0|y)p(y)}{p(x_1 = 0)} \tag{5}$$

Let's start by calculating the features distribution given the class and the normalization constant.

$$
\begin{aligned}
&p(x_1 = 0|y = c) = 0.5 \ \forall c \\
&p(x_1 = 0) = \sum_{c'} \pi_{c'} p(x_1 = 0|y = c') = 0.5 \sum_{c'} \pi_{c'} = 0.5
\end{aligned}
\tag{6}
$$

Since the class conditional is equal to the normalization constant, the posterior is equal to the prior

$$
p(y|x_1 = 0) = \frac{0.5p(y)}{0.5} = \pi \tag{7}
$$

c)
We will perform the same steps as in a and b. Unfortunatly, this item does not have the same simplifications of item b.

$$
p(y|x_2 = 0) = \frac{p(x_2 = 0|y)p(y)}{p(x_2 = 0)} \tag{8}
$$

Let's calculate the class conditional:

$$
p(x_2 = 0|y) = \frac{1}{\sqrt{2\pi}} exp\left[-\frac{1}{2}\mu_c^2\right] = (0.242; 0.399; 0.242) \tag{9}
$$

With the class conditional and the prior, we can compute the joint distribution:

$$
\begin{aligned}
&p(x_2 = 0, y) = p(x_2 = 0|y)p(y) = (0.121; 0.0997; 0.0605) \\
&p(x_2 = 0) = \sum_{c'} p(x_2 = 0, y = c') = 0.281
\end{aligned}
\tag{10}
$$

$$
p(y|x_2 = 0) = (0.430; 0.355; 0.215) \tag{11}
$$

d)
A very interesting pattern observed in the three items above is that in item b the posterior is equal to the prior ($p(y|x_1 = 0) = \pi$). This reflects in the posteriors of item a and c, making $p(y|x_1 = 0, x_2 = 0) = p(y|x_2 = 0)$.
The reason behind this is that $\theta_c = cte$. For constant θ, we can prove that:

$$
p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{\text{cte } p(y)}{\sum c'\text{cte } p(y = c')} = p(y) \tag{12}
$$

Equation 12 says that if $p(x|y)$ is equal for all classes, then there is no redistribution of the priors probabilities. As a result, in item a the prior is redistributed only due to the normal distribution, making $p(y|x_1 = 0, x_2 = 0) = p(y|x_2 = 0)$

Conclusion

In this exercise, we calculated the posteriors of a discrete classifier using mixed features Naive Bayes. The calculation was straighfoward, using Bayes theorem, Naive Bayes property and the computation of the joint distribution and the normalization factor. One particular pattern is this exercise was that the condtional $x_1|y$ was uninformative, which resulted in $p(y|x_1 = 0, x_2 = 0) = (y|x_2 = 0)$