

CAP6610 Machine Learning Homework 2

Hudanyun Sheng

The MATLAB code used in this report are attached at the end of the report.

(1) The mutual information with the classes is defined as:

$$I(X; Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

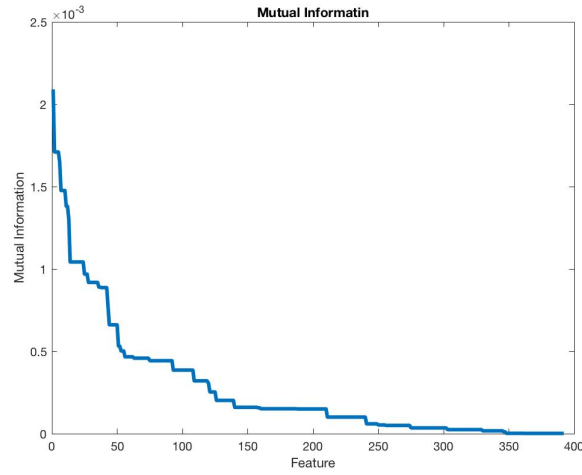


Figure 1: Figure of sorted mutual information for each feature.

Based on the figure, we got to know that the first 120 features after sorting would be the most informative ones. The Naive Bayes Classifier built based on these features would get an error rate of 0.3386.

(2) The generalization of the formula discussed in exercise 3.21 including multinomials is:

$$I_j = \sum_c \sum_j \theta_{jc} \pi_c \log \frac{\theta_{jc}}{\theta_j}$$

where $\sum_c \pi_c = 1$

(3) Suppose the true parameters for the prior Beta distribution are $a = b = 20$, the prior distribution is shown below in figure 2. Now suppose we have the true prior distribution, i.e. the probability of 1 is 0.5, and the probability of 0 is also 0.5.

Now suppose the total number of experiments is 20, the posterior distribution looks like what shown in figure 3.

Now keep the prior beta distribution to be the true one, and suppose the total number of experiments is 200, the posterior distribution is shown in figure 4.

Keep the prior beta distribution to be the true one, and suppose the total number of experiments is 1000, the posterior

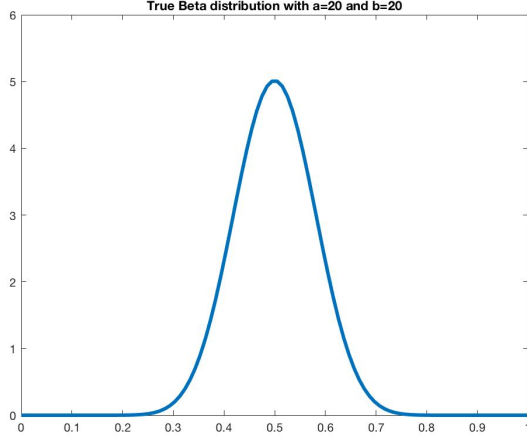


Figure 2: Figure of the beta distribution with $a=20$ and $b=20$

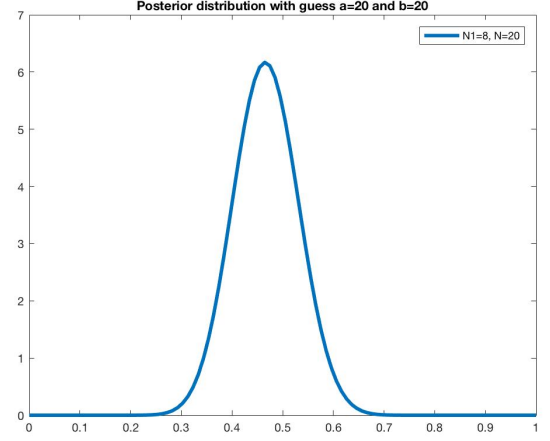


Figure 3: Figure of the posterior distribution with $a=20$, $b=20$, $N_1=8$, $N=20$

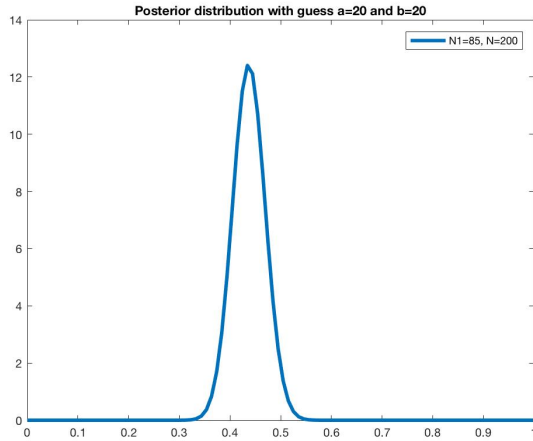


Figure 4: Figure of the posterior distribution with $a=20$, $b=20$, $N_1=85$, $N=200$

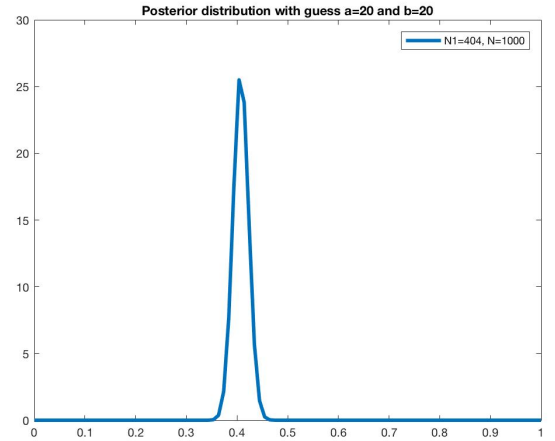


Figure 5: Figure of the posterior distribution with $a=20$, $b=20$, $N_1=404$, $N=1000$

distribution is shown in figure 5.

Now suppose we do not have the true prior distribution, but a prior beta distribution with $a=2$ and $b=4$. The distribution is shown in figure 6. Now suppose the total number of experiments is 20, the posterior distribution is shown in figure 7. Keep the prior beta distribution with $a=2$ and $b=4$, suppose the total number of experiments is 20, the posterior distribution is shown in figure 8.

Keep the prior beta distribution, suppose the total number of experiments is 2000, the posterior distribution is shown in figure 9.

To conclude, when the number of samples is not enough, the prior would take over the posterior distribution, but when

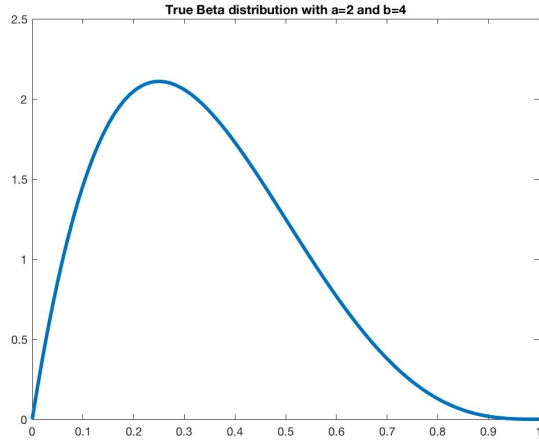


Figure 6: Figure of the beta distribution with $a=2$ and $b=4$

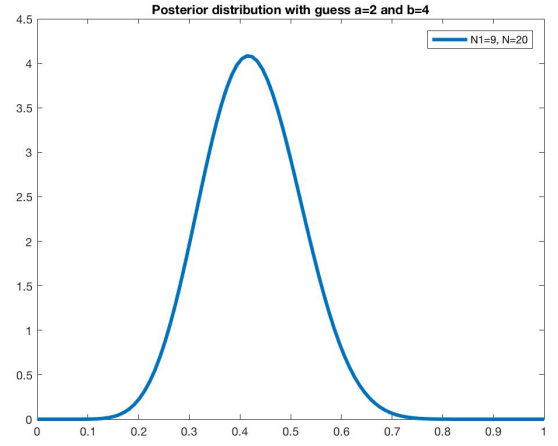


Figure 7: Figure of the posterior distribution with $a=2$, $b=4$, $N_1=9$, $N=20$

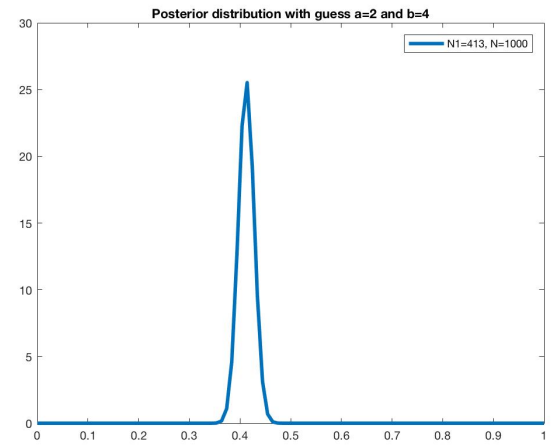
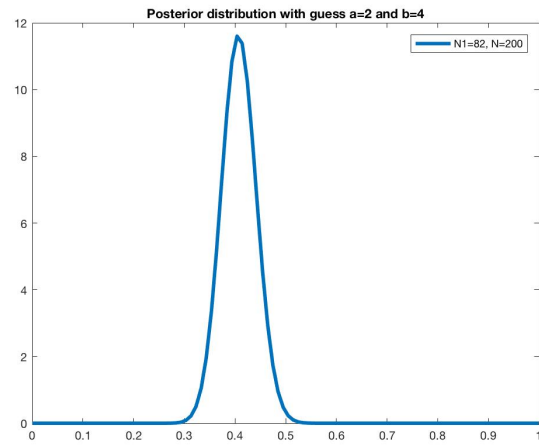


Figure 8: Figure of the beta distribution with $a=2$, $b=4$, $N_1=82$, $N=200$ Figure 9: Figure of the posterior distribution with $a=2$, $b=4$, $N_1=413$, $N=1000$

the number of samples is large enough, the data would take over the posterior distribution.