Homework set 3 TDA231 Algorithms for Machine Learning & Inference

Kevin Jaquier 921119T334
Sandra Viknander 9003012482

March 2018

1 Naive Bayes classifier

For the naive Bayes assumption, in our case, equation 1 holds. For the data that we have from our "happiness" survey we assume it to be Bernoulli distributed since we do have binary data attributes, true or false. As prior distribution for our data we assume it to be Beta distributed since Beta distributions are prior conjugate to Bernoulli distributions. Assuming we have a fairly heterogeneous society with regards to the attributes measured we can assume a fairly uniform distribution between the answers. Thus we can set $\alpha = 1$ and $\beta = 1$ in the prior Beta distributions to have a uniform distribution as prior. To calculate the mean value for our Bernoulli distribution we use equation 2. Since our prior is taken to be uniform, Beta(1,1,x) distributed, our posterior reduces down to equation 1. With equation 2 and our data we can calculate all probabilities for the attributes and the two classes as displayed in table 1. Using a Bayes classifier as in equation 3 we can calculate the probability of a individual being a part of any of the two classes. In the Bayes classifier the prior is taken to be $P(T_{new} = k|X,t) = \frac{N_k}{N}$.

$$P(x_n|t_n = k,X,t) = \prod_{d=1}^{3} P(x_{nd}|t_n,X,t) \tag{1}$$

$$\begin{aligned} k:1 \qquad & p = \frac{1}{n_{trails}} \sum_{i=1}^{n_success} x_i \\ k:0 \qquad & q = 1 - p \end{aligned} \tag{2}$$

$$P(T_{new} = k|x_{new},X,t) = \frac{P(x_{new}|T_{new} = k,X,t)P(T_{new} = k|X,t)}{\sum_{i=1}^{K} P(x_{new}|T_{new} = k_i,X,t)P(T_{new} = k_i|X,t)} \tag{3}$$

Class	Attribute	$P_{Success}$	P_{Fail}
$k=1$			
	Rich	$\frac{3}{4}$	$\frac{1}{4}$
	Married	$\frac{1}{2}$	$\frac{1}{2}$
	Healthy	$\frac{3}{4}$	$\frac{1}{4}$
$k=0$			
	Rich	$\frac{1}{4}$	$\frac{3}{4}$
	Married	$\frac{1}{4}$	$\frac{3}{4}$
	Healthy	$\frac{1}{4}$	$\frac{3}{4}$

Table 1: Table of all the probabilities for the different attributes and classes

1.1 1

Now we can calculate the probability that a person is "married", "Not rich", "Healthy" and "Content". We do this by using our Bayes classifier with the probabilities that we calculated earlier, which gives us equation 4. Thus the probability that a person is "married", "Not rich", "Healthy" and "Content" is $\frac{2}{3}$ according to the survey.

$$\begin{gathered}\frac{\frac{1}{4}\frac{1}{2}\frac{3}{4}1\frac{1}{2}}{\frac{1}{4}\frac{1}{2}\frac{3}{4}1\frac{1}{2}+\frac{3}{4}\frac{1}{4}\frac{1}{4}1\frac{1}{2}}= \\ \frac{\frac{3}{64}}{\frac{3}{64}+\frac{3}{128}}= \\ \frac{6}{9}=\frac{2}{3}\end{gathered} \tag{4}$$

1.2 2

Now calculating the probability that a person is "Not Rich" and "Married" we use the Bayesian classifier as before. But this time we do need to include both cases of "Healthy" and "Not Healthy". We then arrive at equation 5. Thus the new probability is just $\frac{1}{4}$, this is because of the high probability that a person that is "Not Rich", "Married", "Not Healthy" is content is very small.

$$\frac{\frac{1}{4}\frac{1}{2}\frac{3}{4}1\frac{1}{2}+\frac{1}{4}\frac{1}{2}\frac{1}{4}1\frac{1}{2}}{(\frac{1}{4}\frac{1}{2}\frac{3}{4}1\frac{1}{2}+\frac{1}{4}\frac{1}{2}\frac{1}{4}1\frac{1}{2})+(\frac{3}{4}\frac{1}{2}\frac{3}{4}1\frac{1}{2}+\frac{3}{4}\frac{1}{2}\frac{1}{4}1\frac{1}{2})}=\frac{1}{4} \tag{5}$$

2 Extended Naive Bayes

The naive Bayes assumption that we did in the previous question would not work well with our new data set since the first three attributes are dependent on each other. Depending on the size of the data set that we get for these new classes one could do one of two things to mitigate the problem of using the naive Bayes assumption. If our data set is large enough we could treat the first three variables the regular way, i.e. not use the naive Bayes assumption, but use the naive Bayes assumption on the last variable in regards to the three first variable. If the data set is not large enough to confidently treat the three first variables as correlated then grouping them to one variable with values $x \in (0,1,2)$ would be a good candidate as a solution. If one goes with the second option one must also switch to a binomial distribution as likelihood when calculating the posterior distribution. The prior however can still be Beta distributed since it is a prior conjugate to binomial distributions as well.

3 Practical problems

3.1 Bayes Classifier

Our data that we are going to teach our classifier with is in natural numbers, thus we chose a spherical Gaussian distribution for our likelihood function. To calculate the MLE mean value μ_k and the variance σ_k^2 we use equation 6. For our prior we choose also a spherical Gaussian distribution with parameters $\hat{\mu} = 0$ and $\hat{\sigma}^2 = 1$ sins we do not have much prior knowledge about the data. Thus our posterior distribution will have the form as in equation 7. For our classifier we also need to choose a prior that reflects the data set, as is common we use $\frac{N_k}{N}$, as our data set contain a thousand data points for each class the prior for both classes is thus $\frac{1}{2}$. The Bayes classifier can then be derived as in equation 8

$$\begin{aligned}
\mu_k =& \frac{1}{N_k}\sum_{n=1}^{N_k} x_n \\
\sigma_k^2 =& \frac{1}{N_k}\sum_{n=1}^{N_k}(x_n - \mu_k)(x_n - \mu_k)^T
\end{aligned} \tag{6}$$

$$P(x_{new}|T_{new} = k, X, t) = \mathcal{N}(\mu_k, \sigma_k^2\mathbf{I})\mathcal{N}(0,1) \tag{7}$$

$$P(T_{new} = k|x_{new}, X, t) = \frac{\mathcal{N}(\mu_k, \sigma_k^2\mathbf{I})\mathcal{N}(0,1)\frac{1}{2}}{\sum_{i=1}^{K}\mathcal{N}(\mu_i, \sigma_i^2\mathbf{I})\mathcal{N}(0,1)\frac{1}{2})} = \frac{\mathcal{N}(\mu_k, \sigma_k^2\mathbf{I})}{\sum_{i=1}^{K}\mathcal{N}(\mu_i, \sigma_i^2\mathbf{I}))} \tag{8}$$

See the completed Jupyter notebook for the code and results.