CS446: Machine Learning, Fall 2017, Homework 1

Name: Triveni Putti (tputti2)

Worked individually

Problem 1

Solution: (a) Given,

$$Pr[X_i = x|Y = A] = \frac{e^{-\lambda_i^A}(\lambda_i^A)^x}{x!} \quad \text{and} \quad Pr[X_i = x|Y = B] = \frac{e^{-\lambda_i^B}(\lambda_i^B)^x}{x!} \text{ for } i = 1, 2$$

The prior probability $P(Y = A) = \frac{\text{Count of observations with y=A}}{\text{Total observations}}$ and P(Y=B) $=\frac{\text{Count of observations with y=B}}{\text{Total observations}}$.
Since out of the 7 observations, 3 of them have y = A and 4 of them have y = B.

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

Maximum Likelihood Estimate (MLE)is θ that maximizes probability of observed data D

$$\hat{\theta} = argmax_\theta P(D|\theta)$$

So here the likelihood of the dataset will be

$P(X_1 = x_1, X_2 = x_2|Y) = P(X_1 = x_1|Y).P(X_2 = x_2|Y)$(since the features are independent of each other)

$$P(X_1 = x_1, X_2 = x_2|Y = A) = \frac{e^{-\lambda_1^A}(\lambda_1^A)^{x_1}}{x_1!}\frac{e^{-\lambda_2^A}(\lambda_2^A)^{x_2}}{x_2!}$$

$$P(X_1 = x_1, X_2 = x_2|Y = B) = \frac{e^{-\lambda_1^B}(\lambda_1^B)^{x_1}}{x_1!}\frac{e^{-\lambda_2^B}(\lambda_2^B)^{x_2}}{x_2!}$$

Using $P(X_i|\theta) = \theta^{y_i}(1-\theta)^{1-y_i}$, Likelihood (L) can be written as

$$L = (\frac{e^{-\lambda_1^A}(\lambda_1^A)^{x_1}}{x_1!}\frac{e^{-\lambda_2^A}(\lambda_2^A)^{x_2}}{x_2!})^y * (\frac{e^{-\lambda_1^B}(\lambda_1^B)^{x_1}}{x_1!}\frac{e^{-\lambda_2^B}(\lambda_2^B)^{x_2}}{x_2!})^{(1-y)}$$

where if y = 1 then the class will be A and if y = 0 then the class will be B. On simplifying, we get

$$L = (\frac{e^{-\lambda_1^A-\lambda_2^A}(\lambda_1^A)^{x_1}(\lambda_2^A)^{x_2}}{x_1!x_2!})^y * (\frac{e^{-\lambda_1^B-\lambda_2^B}(\lambda_1^B)^{x_1}(\lambda_2^B)^{x_2}}{x_1!x_2!})^{(1-y)}$$

Taking log of this and not considering constants, we get

$$Log(L) = y((-\lambda_1^A-\lambda_2^A)+x_1log(\lambda_1^A)+x_2log(\lambda_2^A))+(1-y)((-\lambda_1^B-\lambda_2^B)+x_1log(\lambda_1^B)+x_2log(\lambda_2^B))$$

For the entire data set, this will be $\sum_{x_1,x_2,y} Log(L)$
To get λ_1^A, we partially differentiate with respect to λ_1^A i.e. we assume other lambdas are constant.The expression becomes

$$= \sum_n (y)(-1 + \frac{x_1}{\lambda_1^A})$$

Since we want to maximize this log likelihood we equate this to zero. Here, y cannot be zero because in that case the class variable will be B (but we are evaluating for A) hence we set,

$$\sum_n (-1 + \frac{x_1}{\lambda_1^A}) = 0$$

Similarly, when we evaluate for other parameters we get similar expressions, which are as follows:

$$\sum_n (-1 + \frac{x_2}{\lambda_2^A}) = 0$$

$$\sum_n (-1 + \frac{x_1}{\lambda_1^B}) = 0$$

$$\sum_n (-1 + \frac{x_2}{\lambda_2^B}) = 0$$

By using the values provided in the data set, we get

$$3 = \frac{9}{\lambda_1^A}$$

$$3 = \frac{18}{\lambda_2^A}$$

$$4 = \frac{20}{\lambda_1^B}$$

$$4 = \frac{16}{\lambda_2^B}$$

The values are $\lambda_1^A = 3, \lambda_2^A = 6, \lambda_1^B = 5, \lambda_2^B = 4$

Solution: (b)

$$\frac{\Pr(X_1 = 2, X_2 = 3 \mid Y = A)}{\Pr(X_1 = 2, X_2 = 3 \mid Y = B)} =?$$

Since the features are independent of each other, this can be written as

$$= \frac{P(X_1 = 2|Y = A).P(X_2 = 3|Y = A)}{P(X_1 = 2|Y = B).P(X_2 = 3|Y = B)}$$

$$= \frac{\frac{e^{-3}(3)^2}{2!}\frac{e^{-6}(6)^3}{3!}}{\frac{e^{-5}(5)^2}{2!}\frac{e^{-4}(4)^3}{3!}} = 1.215$$

Solution: (c) The classifier is

$$h(x) = y_{NB} = argmax_Y P(\mathbf{X}|Y)P(Y) = argmax_Y \prod_{i=1}^{d} P(x_i|Y)P(Y)$$

where d is the number of features. Here, d= 2.

$$= argmax_Y P(x_1|Y)P(x_2|Y)P(Y)$$

i.e. if $P(x_1|Y = A)P(x_2|Y = A)P(Y = A) > P(x_1|Y = B)P(x_2|Y = B)P(Y = B)$ then the classifier predicts A else it predicts B.
Substituting the values, we get

$$= \frac{e^{-3}(3)^{x_1}}{x_1!} \frac{e^{-6}(6)^{x_2}}{x_2!} \frac{3}{7} > \frac{e^{-5}(5)^{x_1}}{x_1!} \frac{e^{-4}(4)^{x_2}}{x_2!} \frac{4}{7}$$

$$= 3^{1+x_1} 6^{x_2} 5^{-x_1} 4^{-1-x_2} > 1$$

If the above condition is satisfied then the predicted class will be A.

Solution: (d) The Naive Bayes classifier in this case is derived as above. Given $\mathbf{x} = [x_1, x_2] = [2, 3]$ Substituting in the final condition in part(c), we get

$$= \frac{27 * 216}{25 * 256} = 0.91125 \text{ which is} < 1$$

Hence the classifier predicts y = B.

Reference - http://www.cs.colostate.edu/ cs545/fall13/dokuwiki/lib/exe/fetch.php?media=wiki

References