CS446: Machine Learning, Fall 2017, Homework 1

Name: Triveni Putti (tputti2)

Worked individually

Problem 3

Solution:
(a) Parametric form of $P(y=1|\mathbf{x}) =?$
Using Bayes theorem, this can be written as

$$= \frac{P(\mathbf{x}|y=1)P(y=1)}{P(\mathbf{x})}$$

Since Y is a binary random variable i.e. y can be either 0 or 1, the denominator can be expanded as follows:

$$= \frac{P(\mathbf{x}|y=1)P(y=1)}{P(\mathbf{x}|y=1)P(y=1) + P(\mathbf{x}|y=0)P(y=0)}$$

Let us call this as equation 1. Assuming that random variables are independent of each other, equation 1 can be further broken down into

$$= \frac{\prod_{i=1}^{n} P(x_i|y=1)P(y=1)}{\prod_{i=1}^{n} P(x_i|y=1)P(y=1) + \prod_{i=1}^{n} P(x_i|y=0)P(y=0)}$$

The above expression will be the parametric form of $P(y=1|\mathbf{x})$

Solution: (b)Dividing both numerator and denominator of equation 1 by$P(x|y=1)P(y=1)$, we get

$$P(y=1|x) = \frac{1}{1 + (\frac{P(\mathbf{x}|y=0)P(y=0)}{P(\mathbf{x}|y=1)P(y=1)})}$$

Now, the second term in the denominator can be expressed as follows:

$$= exp^{ln(\frac{P(\mathbf{x}|y=0)P(y=0)}{P(\mathbf{x}|y=1)P(y=1)})}$$

which can be further simplified to

$$= exp^{ln(P(\mathbf{x}|y=0))+ln(P(y=0))-ln(P(\mathbf{x}|y=1))-ln(P(y=1))}$$

$$= exp^{-(ln(P(\mathbf{x}|y=1))+ln(P(y=1))-ln(P(\mathbf{x}|y=0))-ln(P(y=0)))}$$

Substituting this back we get,

$$\frac{1}{1 + exp^{-(ln(P(\mathbf{x}|y=1))+ln(P(y=1))-ln(P(\mathbf{x}|y=0))-ln(P(y=0)))}}$$

This is now in the form of $\delta(a)$ given in the question, where

$$a = (ln(P(\mathbf{x}|y=1)) + ln(P(y=1)) - ln(P(\mathbf{x}|y=0)) - ln(P(y=0)))$$

Solution: (c) In Naive Bayes algorithm, we assume that the features are conditionally independent given the class label. This allows us to write the class conditional density as a product of one dimensional densities.

$$P(\mathbf{x}|y=c) = \prod_{i=1}^{d} P(x_i|y=c)$$

as there are d features in the data set. The probability density function of a Gaussian distribution for x_i

$$= \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\mu_{ic})^2}{2\sigma_i^2}}$$

Substituting this expression in the above equation, we get

$$P(\mathbf{x}|y=c) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\mu_{ic})^2}{2\sigma_i^2}}$$

Solution: (d) We know from part2 that

$$P(y=1|\mathbf{x}) = \frac{1}{1 + exp^{ln(\frac{P(\mathbf{x}|y=0)P(y=0)}{P(\mathbf{x}|y=1)P(y=1)})}}$$

Let us assume $\frac{P(\mathbf{x}|y=0)P(y=0)}{P(\mathbf{x}|y=1)P(y=1)}$ as 'a',the equation becomes

$$P(y=1|\mathbf{x}) = \frac{1}{1 + exp^{ln(a)}}$$

Now, let us first evaluate a,

$$= \frac{\prod_{i=1}^{d} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\mu_{i0})^2}{2\sigma_i^2}} .P(y=0)}{\prod_{i=1}^{d} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\mu_{i1})^2}{2\sigma_i^2}} .P(y=1)}$$

Note that the variance does not depend on class and hence will be cancelled in the numerator and denominator. Substituting P(y=1) as π and P(y=0) as 1-π and simplifying,

$$a = \prod_{i=1}^{d} e^{\frac{(x_i-\mu_{i1})^2-(x_i-\mu_{i0})^2}{2\sigma_i^2}} * (\frac{1-\pi}{\pi})$$

Now ln(a) will be

$$ln(\prod_{i=1}^{d} e^{\frac{(x_i-\mu_{i1})^2-(x_i-\mu_{i0})^2}{2\sigma_i^2}} * (\frac{1-\pi}{\pi}))$$

$$= \sum_{i=1}^{d} \frac{(x_i-\mu_{i1})^2-(x_i-\mu_{i0})^2}{2\sigma_i^2} + ln(\frac{1-\pi}{\pi})$$

$$= \sum_{i=1}^{d} \frac{\mu_{i1}^2-\mu_{i0}^2-2x_i(\mu_{i1}-\mu_{i0})}{2\sigma_i^2} + ln(\frac{1-\pi}{\pi})$$

$$= \sum_{i=1}^{d} (\frac{\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2} - \frac{x_i(\mu_{i1}-\mu_{i0})}{\sigma_i^2}) + ln(\frac{1-\pi}{\pi})$$

$$= \sum_{i=1}^{d} \frac{x_i(\mu_{i0}-\mu_{i1})}{\sigma_i^2} + \sum_{i=1}^{d} \frac{\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2} + ln(\frac{1-\pi}{\pi})$$

Considering $\frac{(\mu_{i0}-\mu_{i1})}{\sigma_i^2}$ as $-w_i$ and $\sum_{i=1}^{d} \frac{\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2} + ln(\frac{1-\pi}{\pi})$ as $-w_0$, this can be written as

$$= -(w_0 + \sum_{i=1}^{d} w_i x_i)$$

Hence,

$$P(y=1|\mathbf{x}) = \frac{1}{1+exp^{-(w_0+\sum_{i=1}^{d} w_i x_i)}}$$

Therefore, it is proved.

References

Knuth, D. E., Larrabee, T. and Roberts, P. M. (1998). Mathematical writing .