

Assignment 1

Aditya Shinde

August 31, 2017

1 Conditional probability and the chain rule

1

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1)$$

$$P(A|B) \times P(B) = P(A \cap B) \quad (2)$$

$$\text{Similarly, for } P(A \cap B \cap C) \quad (3)$$

$$P(A \cap B \cap C) = P(A \cap (B \cap C)) \quad (4)$$

$$= P(A|B \cap C) \times P(B \cap C) \quad \text{Using (2)} \quad (5)$$

$$= P(A|B \cap C) \times P(B|C) \times P(C) \quad \text{Also using (2)} \quad (6)$$

$$\therefore P(A \cap B \cap C) = P(A|B \cap C) \times P(B|C) \times P(C) \quad (7)$$

2

Consider A and B as two events. Now, the probability that A and B occur at the same time is given by $P(A \cap B)$.

From the law of conditional probability, the probability that A occurs given that B has occurred is,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Similarly, the probability that B occurs given that A has occurred is,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

We can relate both the equations of conditional probability above as,

$$P(A|B) \times P(B) = P(B|A) \times P(A)$$

$$\therefore P(A|B) = P(B|A) \frac{P(A)}{P(B)}$$

2 Total probability

1

Let D_F be the random variable which represents the fair die and D_L be the one which represents the loaded die.

The probability that the fair die will be rolled is p and the probability that the loaded die will be rolled is $(1 - p)$.

The expected value of the die roll will be,

$$\begin{aligned}E[D] &= pE[D_F] + (1 - p)E[D_L] \\&= p\Sigma x_i D_F + (1 - p)\Sigma x_i D_L \\E[D] &= 3.5p + 4.5(1 - p) \\E[D] &= 4.5 - p\end{aligned}$$

2

Variance is,

$$\begin{aligned}Var[D] &= E[D^2] - E[D]^2 \\&= (23.5 - 8.34p) - (4.5 - p)^2 \\&= p^2 - 17.34p + 3.25\end{aligned}$$

3 Naive Bayes

1

$X = \{x_1, x_2, \dots, x_n\}$ and attributes x_2 to x_n are continuous.

So if the continuous attributes are Gaussians, we have to compute,

$$P(x|y) = \frac{1}{\sqrt{2\pi\sigma_{x,y}^2}} e^{-\frac{1}{2}\left(\frac{(x-\mu_{x,y})^2}{\sigma_{x,y}^2}\right)}$$

We have to compute this for $n - 1$ attributes since the first is a boolean. For these $n - 1$ continuous attributes, we have to compute the mean and variance for each which makes $2(n - 1)$ parameters to be estimated. Including the priors for the two target classes, we have to compute $2(2(n - 1) + 1) = 4n - 2$ parameters.

2

If the assumption of conditional independence is not used,

For continuous attributes we have $2(n - 1)$ parameters. For the boolean attribute, there is 1. So the total number parameters to be calculated are,

$$\begin{aligned}&(n - 1)^{1+2(n-1)} - 1 \\&= (n - 1)^{2n-1} - 1\end{aligned}$$

4 Logistic Regression

1

For a X to belong to a particular class, we need

$P(Y = 1|X) > 0.5$ or $P(Y = 0|X) > 0.5$.

But at the decision boundary both classes are equally probable.

Let $h = W^T X$

$$P(Y = 0|X) = P(Y = 1|X)$$

$$\frac{1}{1 + e^{-h}} = \frac{e^{-h}}{1 + e^{-h}}$$

$$1 = e^{-h}$$

$$\ln 1 = \ln e^{-h}$$

$$h = 0$$

$$\therefore W^T X = 0$$

This is the decision boundary. And it is linear in x .

2

The one advantage of logistic regression over naive bayes is that it works well even if the features are not independent. Also logistic regression does well for continous attributes.

The disadvantage is the training process. Since logistic regression uses a gradient based optimization technique, it can get stuck in a local minimum. Also logistic regression is susceptible to overfitting.

5

1

If the Naive Bayes encounters a word which was not observed in training, it will compute its probability as 0.

And since $P(Y|X) = P(Y)\prod P(X_i|Y)$,

$P(Y|X)$ will be computed as 0

This the particular input will be classified as belonging to neither of the classes.

In case of LR, the outcome is the weighted sum of all attributes. So it will output a certain class.