Gender Classification using Bayes Theorem (BEST Example) :
==========

https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781789806076/2/ch02lvl1sec23/gender-classification-bayes-for-continuous-random-variables

Gender classification – Bayes for continuous random variables

So far, we have been given a probability event that belonged to one of a finite number of classes, for example, a temperature was classified as cold, warm, or hot. But how would we calculate the posterior probability if we were given the temperature in °C instead?

For this example, we are given the heights of five men and five women, as shown in the following table:

Height in cm	Gender
180	Male
174	Male
184	Male
168	Male
178	Male
170	Female
164	Female
155	Female
162	Female
166	Female
172	?

Suppose that the next person has a height of 172 cm. What gender is that person more likely to be, and with what probability?

Analysis

One approach to solving this problem could be to assign classes to the numerical values; for example, the people with a height of between 170 cm and 179 cm would be in the same class. With this approach, we may end up with a few classes that are very wide, for example, with a high cm range, or with classes that are more precise but have fewer members, and so we cannot use the full potential of the Bayes classifier. Similarly, using this method, we would not consider the classes of height interval (170, 180) and (180, 190) in centimeters, to be closer to each other than the classes (170, 180) and (190, 200).

Let's remind ourselves of the Bayes' formula here:

$$P(male|height) = P(height|male) * P(male)/P(height)$$

$$= P(height|male) * P(male)/[P(height|male) * P(male) + P(height|female) * P(female)]$$

Expressing the formula in the final form shown here removes the need to normalize the **P(height|male)** and **P(height)** to get the correct probability of a person being male based on their height.

Assuming that the height of people is distributed normally, we could use a normal probability distribution to calculate **P(male|height)**. We assume that **P(male)=0.5**, that is, it is equally likely that the person to be measured is of either gender. A normal probability distribution is determined by the mean, μ, and the variance, σ^2, of the population:

$$f(x|\mu,\sigma^2) = \frac{e\frac{-(x-\mu)^2}{2\sigma^2}}{\sqrt{2\sigma^2\pi}}$$

The following table shows the mean height and variance by gender:

Gender	Mean height	Height variance
Male	176.8	37.2
Female	163.4	30.8

Thus, we could calculate the following:

$$P(height = 172|male) = exp[-(172-176.8)2/(2*37.2)]/[sqrt(2*37.2*\pi)] = 0.04798962999$$

$$P(height = 172|female) = exp[-(172-163.4)2/(2*30.8)]/[sqrt(2*30.8*\pi)] = 0.02163711333$$

Note that these are not the probabilities, just the values of the probability density function. However, from these values, we can already observe that a person with a height of 172 cm is more likely to be male than female because, as follows:

$$P(height = 172|male) > P(height = 172|female)$$

To be more precise:

$$P(male|height = 172)$$
$$= P(height = 172|male) * P(male)/[P(height = 172|male) * P(male) + P(height = 172|female) * P(female)]$$

$$= 0.04798962999 * 0.5/[0.04798962999 * 0.5 + 0.02163711333 * 0.5] = 0.68924134178 \sim 68.9\%$$

Therefore, the probability of a person with a height of **172** cm being male is of **68.9%**.