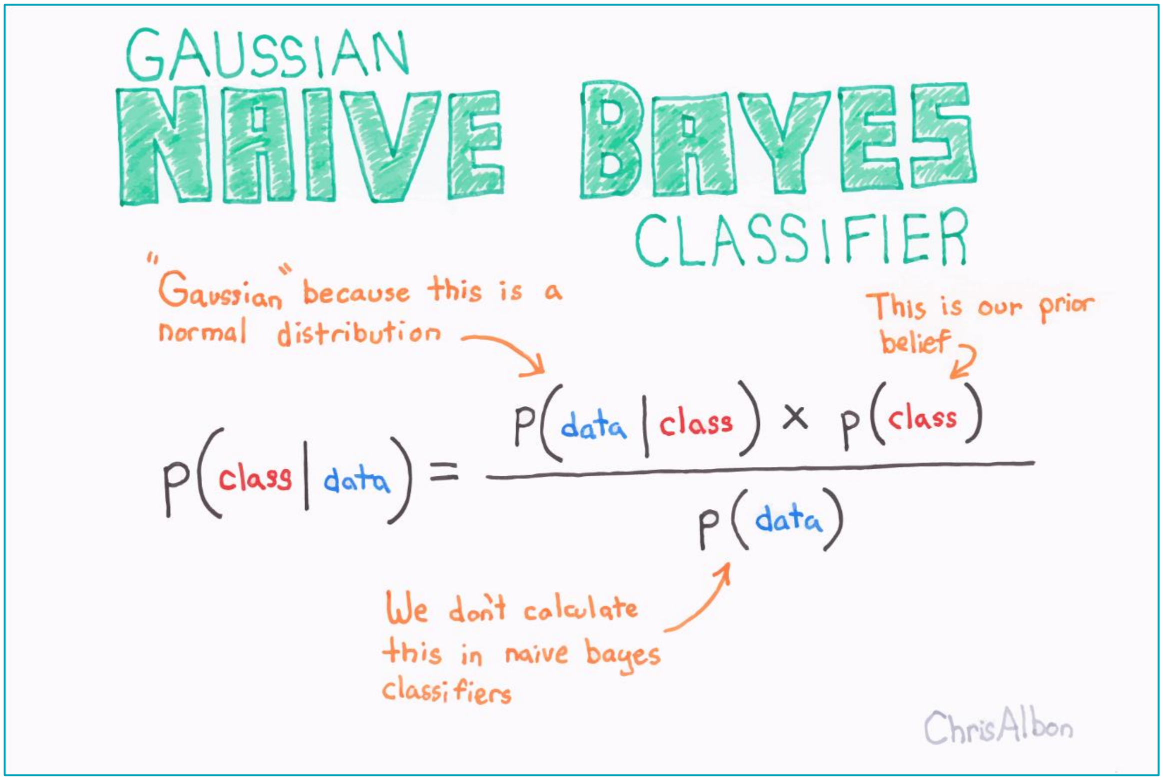

Naïve Bayes Classification

PART C

03717357 | COMP6930 | 24-04-18

Part C - Manual Calculation (10 marks)

Consider the following table that contains data on the measurements of several flowers of two different species. Assuming that Petal Length and Petal Width are normally distributed with respect to the species, manually classify a flower of Petal Length = 5.2 cm and Petal Width = 2.3 cm. Show your working a *.pdf* file called *PartC.pdf*

Petal Length (Cm)	Petal Width (Cm)	Species
1.4	0.2	0
1.4	0.2	0
1.3	0.2	0
1.5	0.2	0
1.4	0.2	0
1.7	0.4	0
5.7	2.3	1
4.9	2.0	1
6.7	2.0	1
4.9	1.8	1
5.7	2.1	1
6.0	1.8	1

Let's Calculate the Priors.

$$P(species = 0) = P(species = 1) = \frac{6}{12} = 0.5$$

Now Petal Width and Petal Length are distributed normally with respect to Species.
Let's use the estimates of the mean and variance to describe these normal distributions.

$$\mu_{petal\ length\ |\ species=0} = \frac{\sum_i^n petal\ length_{i|species=0}}{n} = \frac{8.7}{6} = 1.45$$

$$\sigma^2_{petal\ length\ |\ species=0} = \frac{\sum_i^n (petal\ length_{i|species=0} - \mu_{petal\ length\ |\ species=0})^2}{n-1} = 0.019$$

$$\mu_{petal\ length\ |\ species=1} = \frac{\sum_i^n petal\ length_{i|species=1}}{n} = \frac{33.9}{6} = 5.65$$

$$\sigma^2_{petal\ length\ |\ species=1} = \frac{\sum_i^n (petal\ length_{i|species=1} - \mu_{petal\ length\ |\ species=1})^2}{n-1} = 0.471$$

$$\mu_{petal\ width\ |\ species=0} = \frac{\sum_i^n petal\ width_{i|species=0}}{n} = \frac{1.4}{6} = 0.233$$

$$\sigma^2_{petal\ width\ |\ species=0} = \frac{\sum_i^n (petal\ width_{i|species=0} - \mu_{petal\ width\ |\ species=0}\ )^{\ 2}}{n-1} = 0.0066$$

$$\mu_{petal\ width\ |\ species=1} = \frac{\sum_i^n petal\ width_{i|species=1}}{n} = \frac{12}{6} = 2.0$$

$$\sigma^2_{petal\ width\ |\ species=1} = \frac{\sum_i^n (petal\ width_{i|species=1} - \mu_{petal\ width\ |\ species=01})^{\ 2}}{n-1} = 0.0036$$

Now the probability density function of the Normal distribution is given by :

$$f(x\colon \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Given a sample x= {petal length=5.2, petal width=2.3}, we wish to establish which Species Class has a higher probability. I.e.

$$argmax\{P(species = 0\mid x), P(species = 1\mid x\}$$

Now

$$P(species = 0|x) = \frac{P(species=0)P(petal\ length=5.2|species=0)P(petal\ width=2.3|species=0)}{evidence}$$

$$P(species = 1|x) = \frac{P(species=1)P(petal\ length=5.2|species=1)P(petal\ width=2.3|species=1)}{evidence}$$

So
$argmax\{P(species = 0\mid x), P(species = 1\mid x)\}\ \propto$

$$Max\{P(species = 0)P(petal\ length = 5.2|species = 0)P(petal\ width = 2.3|species = 0),$$
$$P(species = 1)P(petal\ length = 5.2|species = 1)P(petal\ width = 2.3|species = 1)\}$$

Let's calculate some of the probability densities:

$$P(petal\ length = 5.2|species = 0)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2_{petal\ length\ |\ species=0}}} e^{\frac{-(5.2-\mu_{petal\ length\ |\ species=0})^2}{2\sigma^2_{petal\ length\ |\ species=0}}}$$
$$= \frac{1}{\sqrt{2\pi(0.019)}} e^{\frac{-(5.2-1.45)^2}{2*0.019}} = 0$$

$$P(petal\ length = 5.2|species = 1) = \frac{1}{\sqrt{2\pi\sigma^2_{petal\ length\ |\ species=1}}} e^{\frac{-(5.2-\mu_{petal\ length\ |\ species=1})^2}{2\sigma^2_{petal\ length\ |\ species=1}}}$$
$$= \frac{1}{\sqrt{2\pi(0.471)}} e^{\frac{-(5.2-5.65)^2}{2*0.471}} = 0.5812 * 0.8066 = 0.4689$$

$$P(petal\ width = 2.3|species = 1) = \frac{1}{\sqrt{2\pi\sigma^2_{petal\ width\ |\ species=1}}} e^{\frac{-(5.2-\mu_{petal\ width\ |\ species=1})^2}{2\sigma^2_{petal\ width\ |\ species=1}}}$$
$$= \frac{1}{\sqrt{2\pi(0.0036)}} e^{\frac{-(2.3-2.0)^2}{2*0.0036}} = 2.477 * \ 10^{-5}$$

So without going further we can conclude that:

$$P(species = 1)P(petal\ length = 5.2|species = 1)P(petal\ width = 2.3|species = 1) >$$
$$P(species = 0)P(petal\ length = 5.2|species = 0)P(petal\ width = 2.3|species = 0)$$

Since

$$P(petal\ length = 5.2|species = 0) = 0$$

And thus we classify sample x= {petal length=5.2, petal width=2.3} as having species = 1