Tutorial solutions week 12

1. (a) What are the prior probabilities for each sport, ie $P(basketball)$, $P(football)$ and $P(cricket)$?

The prior probabilities for each class are estimated by counting the instances:

$$P(basketball) = \frac{5}{15}$$

$$P(football) = \frac{5}{15}$$

$$P(cricket) = \frac{5}{15}$$

1. (b) What are $P(height \mid basketball)$, $P(height \mid football)$ and $(height \mid cricket)$?

Statistically speaking $P(height \mid basketball)$ is the distribution of the height attribute among the instances of the basketball class.

Probabilistically speaking this is the probability density function which tells us how likely it is for attribute height to equal to a specific value, if that instance is a basketball instance. For example, $P(200 \mid basketball)$ is the probability that a basketball player has a height of 200 cm. If enough instances have been collected it will approximate to a normal distribution. Same applies to $P(height \mid football)$ and $P(height \mid cricket)$.

1. (c) Using Bayes rule, find the posterior probabilities $P(basketball \mid height)$, $P(basketball \mid height)$ and $P(basketball \mid height)$.

$P(basketball \mid height) = \frac{P(height \mid basketball) \times P(basketball)}{P(height)}$, where

$$P(height) = P(height \mid basketball) \times P(basketball) + P(height \mid football) \times P(football) + P(height \mid cricket) \times P(cricket)$$

Similarly:

$$P(football \mid height) = \frac{P(height \mid football) \times P(football)}{P(height)}$$

$$P(cricket \mid height) = \frac{P(height \mid cricket) \times P(cricket)}{P(height)}$$

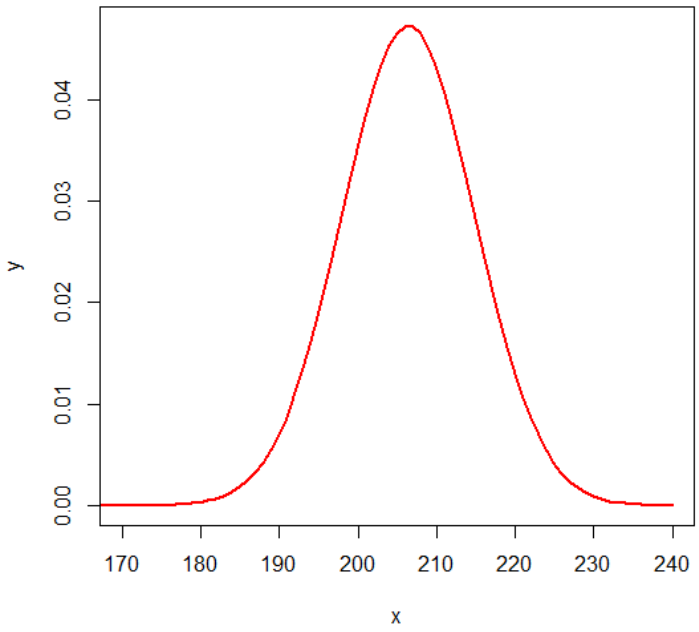
Note that the denominator is always the same.

1. (d) and 1. (e) Sketch the posterior probabilities

Since we can ignore the denominator, the posterior probability $P(basketball \mid height) \propto P(height \mid basketball) \times P(basketball)$.

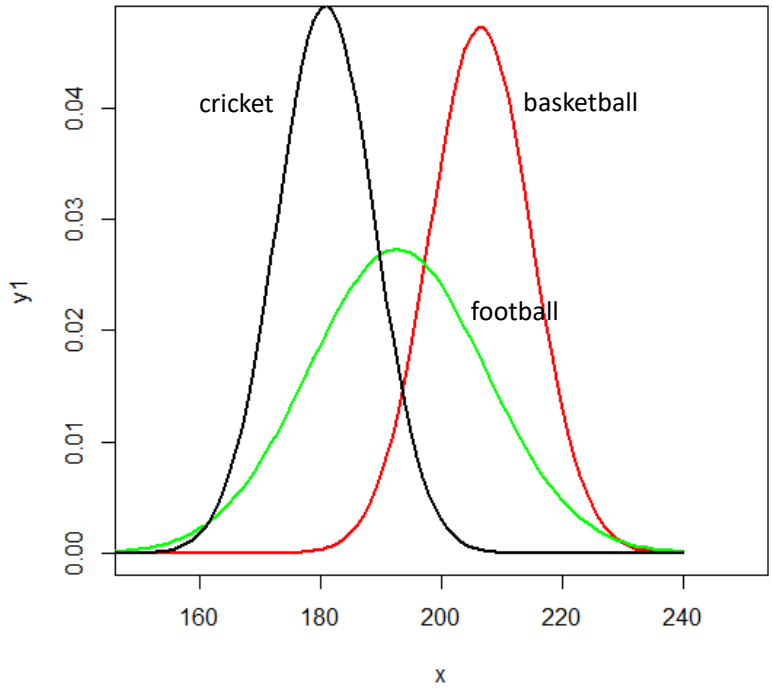
$P(basketball)$ is the prior probability of the basketball class and given the sample set this is already known. Therefore, to sketch the posterior probability we just need to work on $P(height \mid basketball)$. As we've discussed in 1 b) this is a normal distribution. Therefore we just need to get the **mean** and **standard deviation** and then we'll conveniently get the graph.

The basketball sample set: {214, 200, 213, 210, 195}
Mean(basketball)=206.4
Standard deviation(basketball)=8.44



When sketching the graph manually, apply the 68% - 95% - 99.7% Rule:
The area within 1 standard deviation (206.4 ∓ 8.44) is around **68%** of the total area.
The area within 2 standard deviations (206.4 ∓ 16.88) is around **95%.**
The area within3 standard deviations is around **99.7%.**

Similarly, we can get the posteriors for football and cricket.
Mean(football)=192.6,
Standard deviation (football)=14.62,
Mean(cricket)=180.8,
Standard deviation (cricket)=8.11



1. (f) How would a person who is 190 cm tall be classified?

We apply the probability density function: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

$P(basketball \mid height) \propto P(height \mid basketball) \times P(basketball)$ and

$P(190 \mid basketball) = \frac{1}{8.44\sqrt{2\pi}} e^{-\frac{(190-206.4)^2}{2(8.44)^2}} \approx 0.006$

$P(basketball) \approx 0.333$

$$\boldsymbol{P(basketball \mid 190) \propto 0.006 \times 0.333}$$
$$\boldsymbol{\approx 0.002}$$

$P(football \mid height) \propto P(height \mid football) \times P(football)$ and

$P(190 \mid football) = \frac{1}{14.62\sqrt{2\pi}} e^{-\frac{(190-192.6)^2}{2(14.62)^2}} \approx 0.025$

$P(football) \approx 0.333$

$$\boldsymbol{P(football \mid 190) \propto 0.025 \times 0.333}$$
$$\boldsymbol{\approx 0.008}$$

$P(cricket \mid height) \propto P(height \mid cricket) \times P(cricket)$ and

$P(190 \mid cricket) = \frac{1}{8.11\sqrt{2\pi}} e^{-\frac{(190-180.8)^2}{2(8.11)^2}} \approx 0.020$

$P(cricket) \approx 0.333$

$$\boldsymbol{P(cricket \mid 190) \propto 0.020 \times 0.333}$$
$$\boldsymbol{\approx 0.007}$$

As $P(football \mid 190) = 0.008 > P(cricket \mid 190) = 0.007 > P(basketball \mid 190) = 0.002$ is the most likely class is **football** as it has the highest probability. We therefore classify the person as a **football player**.

1. (g) Work out an approximate error rate.

The error rate can be estimated by calculating the area of the overlapped parts under the curves. Note that in this graph of three curves, there is an area covered by all three curves, and this area will need to be added more than once. The total area under each curve is 1, so divide the total area of the overlapped parts by 3.

$$Error = \frac{\big(Area(basketball) \cap Area(football)\big) + \big(Area(basketball) \cap Area(cricket)\big) + (Area(cricket) \cap Area(football))}{Area(basketball) + Area(football) + Area(cricket)}$$

This is around **0.4 to 0.5.**

1. (h) Is this a good classifier?

This is not a good classifier, as the overlapping area is large, implying a high error rate.

1. (i) Could the classifier be improved by getting the heights of 1000 sportspeople?

It depends. In general when more instances are collected the estimation of the distributions will be more accurate, so it is possible that the curves will be more separated from each other; but if the distributions of height in these classes are indeed similar then there will still be a large overlapping area, and in that case we need to find better features.

2. Suppose we want a classifier for 'Flu' and 'Well'. We find 100 people who are well and 100 who have the flu and build a classifier based on the univariate normal distribution.

2. (a) What is wrong with this procedure?

By doing so we are assuming that there are as many people with measles as are well, which is not true in the real world. In other words, the sample distribution should be like the distribution in the whole population, but arbitrarily sampling 100 measles and 100 well is equivalent to saying $\mathrm{P}(\mathrm{measles}) = \mathrm{P}(\mathrm{well})$, and that is not true of the population.

2. (b) What if we built a decision tree?

Decision trees also bring the prior probabilities into consideration, though in a different way, when calculating the information gain/information ratio. So, this procedure is wrong as well.

3. Consider the following data from a factory:

Run	Operator	Machine	Length	Overtime	output
1	Joe	a	51	no	high
2	Sam	b	85	yes	low
3	Jim	b	63	no	low
4	Jim	b	39	no	high
5	Joe	c	32	no	high
6	Sam	c	47	no	low
7	Joe	c	63	no	low
8	Jim	a	70	yes	low
9	Jim	a	51	yes	??

3. (a) Using only the nominal attributes of runs 1 to 8, how will case 9 be classified by a naive Bayes classifier.

$P(high \mid Jim, a, yes) \propto P(Jim \mid high) \times P(a \mid high) \times P(yes \mid high) \times P(high)$

$P(Jim \mid high)$ can be estimated by counting the instances. Among all high instances, 1 out of 3 has Jim as operator, so:

$P(Jim \mid high) = \frac{1}{3}$

Similarly,

$P(high \mid Jim, a, yes) \propto \frac{1}{3} \times \frac{1}{3} \times 0 \times \frac{3}{8} = 0$

$P(low \mid Jim, a, yes) \propto P(Jim \mid low) \times P(a \mid low) \times P(yes \mid low) \times P(low)$

$P(low \mid Jim, a, yes) \propto \frac{2}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{5}{8} = 0.02$

Since $\mathbf{P(high \mid Jim, a, yes) < P(low \mid Jim, a, yes)}$, we classify the instance as low.

3. (b) How would case 9 be classified by naive Bayes if all attributes are used?

If all attributes are used, we need to calculate $P(high \mid Jim, a, 51, yes)$ and $P(low \mid Jim, a, 51, yes)$ then compare them:

$P(high \mid Jim, a, 51, yes) \propto P(Jim|high) \times P(a|high) \times P(51|high) \times P(yes|high) \times P(high)$

$P(low \mid Jim, a, 51, yes) \propto P(Jim \mid low) \times P(a \mid low) \times P(51 \mid low) \times P(yes \mid low) \times P(low)$

For instances where $operator = Jim$ and $output = high$ the values for $length$ are $(85,63,47,63,70)$ with a $mean = 65.60$ and $standard\ deviation = 12.28$.

$P(51 \mid high) = \frac{1}{12.28\sqrt{2\pi}} e^{-\frac{(51-65.60)^2}{2(12.28)^2}} \approx 0.016$

For instances where $operator = Jim$ and $output = low$ the values for $length$ are $(51,39,32)$ with a $mean = 40.66$ and $standard\ deviation = 7.84$.

$$P(51 \mid low) = \frac{1}{7.84\sqrt{2\pi}} e^{-\frac{(51-40.66)^2}{2(7.84)^2}} \approx 0.021$$

Therefore:

$$P(high \mid Jim, a, 51, yes) \propto \frac{1}{3} \times \frac{1}{3} \times 0.016 \times 0 \times \frac{3}{8} = 0$$

$$P(low \mid Jim, a, 51, yes) \propto \frac{2}{5} \times \frac{1}{5} \times 0.021 \times \frac{2}{5} \times \frac{5}{8} = 0.00042$$

The classification result is the same as the previous question: low.

3. (c) How does the result of the previous question change if the Laplace correction is used?

With Laplace correction we add one to all counts. We start with the same functions:

$P(high \mid Jim, a, 51, yes) \propto P(Jim|high) \times P(a|high) \times P(51|high) \times P(yes|high) \times P(high)$

$P(low \mid Jim, a, 51, yes) \propto P(Jim \mid low) \times P(a \mid low) \times P(51 \mid low) \times P(yes \mid low) \times P(low)$

Calculations for output=high:

$$P(Jim \mid high) = \frac{count(Jim = high) + 1}{(count(Joe = high) + 1) + (count(sam = high) + 1) + (count(jim = high) + 1)}$$

$$= \frac{1+1}{3+1+2}$$

$$= \frac{1}{3}$$

$$P(a \mid high) = \frac{1+1}{(1+1)+(1+1)+(1+1)}$$

$$= \frac{1}{3}$$

$P(51 \mid high) = 0.016$ (from the calculation in question 3. b). Laplace correction is only used for nominal attributes, so this value remains the same as calculated previously.

$$P(yes \mid high) = \frac{0+1}{(0+1)+(3+1)}$$

$$= \frac{1}{5}$$

$$P(high) = \frac{3+1}{(3+1)+(5+1)}$$

$$= \frac{4}{10}$$

Calculations for output=low:

$$P(Jim \mid low) = \frac{count(Jim = low) + 1}{(count(Joe = low) + 1) + (count(sam = low) + 1) + (count(jim = low) + 1)}$$

$$= \frac{3}{2+3+3}$$

$$= \frac{3}{8}$$

$$P(a \mid low) = \frac{1+1}{(1+1)+(2+1)+(2+1)}$$

$$= \frac{1}{4}$$

$P(51 \mid low) = 0.021$ (from the calculation in question 3. b). Laplace correction is only used for nominal attributes, so this value remains the same as calculated previously.

$$P(yes \mid low) = \frac{2+1}{(2+1)+(3+1)}$$

$$= \frac{3}{7}$$

$$P(low) = \frac{5+1}{(3+1)+(5+1)}$$

$$= \frac{6}{10}$$

Therefore:

$P(high \mid Jim, a, 51, yes) = \frac{1}{3} \times \frac{1}{3} \times 0.016 \times \frac{1}{5} \times \frac{4}{10} = 0.000142$, and

$P(low \mid Jim, a, 51, yes) = \frac{3}{8} \times \frac{1}{4} \times 0.021 \times \frac{3}{7} \times \frac{6}{10} = 0.000506$

The classification result remains the same as the previous question: low.