1 Probability and Distributions

1) Consider the following registry where an experiment is repeated six times and four events (A, B, C and D) are detected. Considering frequentist estimates for probabilities, compute:

	D	C	B	A
1	0	0	1	1
2	0	1	1	1
3	1	0	0	0
4	1	0	0	0
5	0	0	0	0
6	0	0	0	0

- $p(A)$
- $p(A, B)$
- $p(B \mid A)$
- $p(A, B, C)$
- $p(A \mid B, C)$
- $p(A, B, C, D)$
- $p(D \mid A, B, C)$

This one is not rocket science, just calculate the probabilities normally and then use probability rules to compute the rest.

1

$$\#\{D\} = 2 \; \#\{C\} = 1 \; \#\{B\} = 2 \; \#\{A\} = 2$$

$$p(D) = \tfrac{2}{6} \; p(C) = \tfrac{1}{6} \; p(B) = \tfrac{2}{6} \; p(A) = \tfrac{2}{6}$$

2

$$\#\{A, B\} = 2 \; \#\{A, D\} = 0 \; \#\{B, D\} = 0$$
$$\#\{A, C\} = 1 \; \#\{B, C\} = 1 \; \#\{C, D\} = 0$$

$$p(A, B) = \tfrac{2}{6} \; p(A, D) = 0 \; p(B, D) = 0$$
$$p(A, C) = \tfrac{1}{6} \; p(B, C) = \tfrac{1}{6} \; p(C, D) = 0$$

3

$$\#\{A, B, C\} = 1 \; \#\{A, B, D\} = 0 \; \#\{A, C, D\} = 0 \; \#\{B, C, D\} = 0$$

$$p(A, B, C) = \tfrac{1}{6} \; p(A, B, D) = 0 \; p(A, C, D) = 0 \; p(B, C, D) = 0$$

4

$$\#\{A, B, C, D\} = 0 \; p(A, B, C, D) = 0$$

- $p(A) = \frac{2}{6}$
- $p(A, B) = \frac{2}{6}$
- $p(B \mid A) = \frac{p(A,B)}{p(A)} = \frac{\frac{2}{6}}{\frac{2}{6}} = 1$
- $p(A, B, C) = \frac{1}{6}$
- $p(A \mid B, C) = \frac{p(A,B,C)}{p(B,C)} = \frac{\frac{1}{6}}{\frac{1}{6}} = 1$
- $p(A, B, C, D) = 0$
- $p(D \mid A, B, C) = \frac{p(A,B,C,D)}{p(A,B,C)} = \frac{0}{\frac{1}{6}} = 0$

2) Consider the following set of height measures in centimeters of a group of people:

$$X \,|\, 180 \; 160 \; 200 \; 171 \; 159 \; 150$$

What are the maximum likelihood parameters of a gaussian distribution for this set of points? Plot it approximately.

Just compute expected value and variance and insert into Gaussian Distribution expression.

The maximum likelihood gaussian is defined by the sample mean and standard deviation. Let us compute them:

$$\mu = \frac{180 + 160 + 200 + 171 + 159 + 150}{6} = 170$$

$$\sigma = \left(\frac{1}{6-1}\left((180-170)^2 + (160-170)^2 + \right.\right.$$
$$\left.\left. + (200-170)^2 + (171-170)^2 + (159-170)^2 + (150-170)^2\right)\right)^{\frac{1}{2}} = 18.0111$$

Having the parameters, we can write the expression:

$$N(x \mid \mu, \sigma) = \frac{1}{18.0111\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-170}{18.0111}\right)^2\right)$$

3) Consider the following set of two dimensional measures:

X_1	-2	-1	0	-2
X_2	2	3	1	1

#observations $= m = 4$

What are the maximum likelihood parameters of a Gaussian distribution for this set of points? What is the shape of the Gaussian? Draw it approximately using a contour map.

Same scheiße, but with more dimensions. Now we need the <u>vector of expected values</u> and the <u>covariance matrix</u>.

The maximum likelihood gaussian is defined by the sample mean vector and the covariance matrix. Let us compute them:

$$\mu = \frac{1}{4}\left(\begin{bmatrix}-2\\2\end{bmatrix}+\begin{bmatrix}-1\\3\end{bmatrix}+\begin{bmatrix}0\\1\end{bmatrix}+\begin{bmatrix}-2\\1\end{bmatrix}\right)=\begin{bmatrix}-1.25\\1.75\end{bmatrix}$$

$\mu : \mu_i = E[X_i]$

$$\Sigma_{00} = \frac{1}{4-1}[(-2+1.25)(-2+1.25)+(-1+1.25)(-1+1.25)+ \\ +(0+1.25)(0+1.25)+(-2+1.25)(-2+1.25)] \approx 0.9167$$

$$\Sigma_{01} = \frac{1}{4-1}[(-2+1.25)(2-1.75)+(-1+1.25)(3-1.75)+ \\ +(0+1.25)(1-1.75)+(-2+1.25)(1-1.75)] \approx -0.0833$$

$$\frac{1}{m-1}\left[(x_1(0)-E(x_1))(x_2(0)-E(x_2)) + (x_1(1)-E(x_1))(x_2(1)-E(x_2))+\ldots\right]$$

$$\Sigma_{10} = \frac{1}{4-1}[(2-1.75)(-2+1.25)+(3-1.75)(-1+1.25)+ \\ +(1-1.75)(0+1.25)+(1-1.75)(-2+1.25)] \approx -0.0833$$

$$\Sigma_{11} = \frac{1}{4-1}[(2-1.75)(2-1.75)+(3-1.75)(3-1.75)+ \\ +(1-1.75)(1-1.75)+(1-1.75)(1-1.75)] \approx 0.9167$$

- $\Sigma : \Sigma_{(i-1)(j-1)} = Cov(X_i, X_j)$
  $= E((X_i - E(X_i))(X_j - E(X_j)))$
- $\Sigma_{ij} = \Sigma_{ji}$ (Σ is symmetric)

Then, just compute the multivariate gaussian:

To compute the expression for the multivariate gaussian we need to compute the determinant of Σ and its inverse:

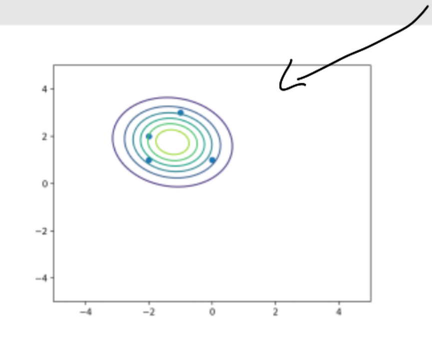
$$det(\Sigma) = det\begin{pmatrix}0.9167 & -0.0833\\ -0.0833 & 0.9167\end{pmatrix} = (0.9167\cdot 0.9167)-(-0.0833\cdot -0.0833) = 0.8333$$

$$\Sigma^{-1} = \frac{1}{0.8333}\begin{bmatrix}0.9167 & 0.0833\\ 0.0833 & 0.9167\end{bmatrix} = \begin{bmatrix}1.1 & 0.1\\ 0.1 & 1.1\end{bmatrix}$$

So, we can write the expression for a two dimensional input $\mathbf{x} = [x_0\ x_1]^T$ as follows.

$$f_{\mathbf{X}}(x_1,\ldots,x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}$$

$$N(\mathbf{x}\mid\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{2}{2}}\sqrt{0.8333}}\exp\left(-\frac{1}{2}\left(\begin{bmatrix}x_0\\x_1\end{bmatrix}-\begin{bmatrix}-1.25\\1.75\end{bmatrix}\right)^T\begin{bmatrix}1.1 & 0.1\\0.1 & 1.1\end{bmatrix}\left(\begin{bmatrix}x_0\\x_1\end{bmatrix}-\begin{bmatrix}-1.25\\1.75\end{bmatrix}\right)\right)$$



(I have no idea how to estimate the plot of this graph)

2 Simple Bayesian Learning

2) From the following training set:

X_1	X_2	X_3	X_4	X_5	C
1	1	0	1	0	a
1	0	0	1	1	a
1	0	0	1	1	a
1	1	1	0	1	b
0	0	1	1	1	b
1	0	0	0	0	c

a) Compute the class for the pattern $\mathbf{x} = [1\,0\,1\,0\,1]^T$ under the Naive Bayes assumption.

Estimate priors:

X_1	X_2	X_3	X_4	X_5	C	
1	1	0	1	0	a	
1	0	0	1	1	a	3
1	0	0	1	1	a	
1	1	1	0	1	b	2
0	0	1	1	1	b	
1	0	0	0	0	c	1

$$p(C=a) = \frac{3}{6}$$

$$p(C=b) = \frac{2}{6}$$

$$p(C=c) = \frac{1}{6}$$

Estimate likelihood for each case:

X_1	X_2	X_3	X_4	X_5	C
1	1	0	1	0	a
1	0	0	1	1	a
1	0	0	1	1	a
1	1	1	0	1	b
0	0	1	1	1	b
1	0	0	0	0	c

Take for example the first case:

X_1	$p(X_1 \mid C=a)$	$p(X_1 \mid C=b)$	$p(X_1 \mid C=c)$
0	$\frac{0}{3} \to 0$	$\frac{1}{2}$	$\frac{0}{1}$
1	$\frac{3}{3} \to 1$	$\frac{1}{2}$	$\frac{1}{1}$

X_2	$p(X_2 \mid C=a)$	$p(X_2 \mid C=b)$	$p(X_2 \mid C=c)$
0	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{1}$
1	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{0}{1}$

X_3	$p(X_3 \mid C=a)$	$p(X_3 \mid C=b)$	$p(X_3 \mid C=c)$
0	$\frac{3}{3}$	$\frac{0}{2}$	$\frac{1}{1}$
1	$\frac{0}{3}$	$\frac{2}{2}$	$\frac{0}{1}$

X_4	$p(X_4 \mid C=a)$	$p(X_4 \mid C=b)$	$p(X_4 \mid C=c)$
0	$\frac{0}{3}$	$\frac{1}{2}$	$\frac{1}{1}$
1	$\frac{3}{3}$	$\frac{1}{2}$	$\frac{0}{1}$

X_5	$p(X_5 \mid C=a)$	$p(X_5 \mid C=b)$	$p(X_5 \mid C=c)$
0	$\frac{1}{3}$	$\frac{0}{2}$	$\frac{1}{1}$
1	$\frac{2}{3}$	$\frac{2}{2}$	$\frac{0}{1}$

Then, compute the posterior for each class with the query vector and the Bayes theorem:

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

$$p(C=a \mid X_1=1, X_2=0, X_3=1, X_4=0, X_5=1)$$

$$= \frac{p(C=a)\,p(X_1=1, X_2=0, X_3=1, X_4=0, X_5=1 \mid C=a)}{p(X_1=1, X_2=0, X_3=1, X_4=0, X_5=1)}$$

These terms are 0

$$= \frac{p(C=a)\,p(X_1=1 \mid C=a)\,p(X_2=0 \mid C=a)\,\overbrace{p(X_3=1 \mid C=a)\,p(X_4=0 \mid C=a)}\,p(X_5=1 \mid C=a)}{p(X_1=1, X_2=0, X_3=1, X_4=0, X_5=1)}$$

$$= 0$$

$$p\left(C=b \mid X_1=1, X_2=0, X_3=1, X_4=0, X_5=1\right)$$

$$= \frac{p(C=b)\, p\left(X_1=1, X_2=0, X_3=1, X_4=0, X_5=1 \mid C=b\right)}{p\left(X_1=1, X_2=0, X_3=1, X_4=0, X_5=1\right)}$$

$$= \frac{p(C=b)\, p(X_1=1 \mid C=b)\, p(X_2=0 \mid C=b)\, p(X_3=1 \mid C=b)\, p(X_4=0 \mid C=b)\, p(X_5=1 \mid C=b)}{p\left(X_1=1, X_2=0, X_3=1, X_4=0, X_5=1\right)}$$

$\neq 0$ As we will see later, as this is the only non-zero term, this will be the output of the classifier.

$$p\left(C=c \mid X_1=1, X_2=0, X_3=1, X_4=0, X_5=1\right)$$

$$= \frac{p(C=c)\, p\left(X_1=1, X_2=0, X_3=1, X_4=0, X_5=1 \mid C=c\right)}{p\left(X_1=1, X_2=0, X_3=1, X_4=0, X_5=1\right)}$$

$$= \frac{p(C=c)\, p(X_1=1 \mid C=c)\, p(X_2=0 \mid C=c)\, \overbrace{p(X_3=1 \mid C=c)}^{0}\, p(X_4=0 \mid C=c)\, \overbrace{p(X_5=1 \mid C=c)}^{1}}{p\left(X_1=1, X_2=0, X_3=1, X_4=0, X_5=1\right)}$$

$= 0$

b) What is the posterior probability $p(b \mid \mathbf{x})$? Normally, we would calculate the term in the denominator of the posteriors and use it to calculate the asked probability.

$$\sum_{y \in \{a,b,c\}} p\left(C=y \mid X_1=1, X_2=0, X_3=1, X_4=0, X_5=1\right)=1$$

↳ Terms for $y=a$ and $y=c$ are 0.

$$\cancel{P(C=a \mid x)} + P(C=b \mid x) + \cancel{P(C=c \mid x)} = 1$$

$$\Leftrightarrow P(C=b \mid x) = 1$$

c) What do you do if we have missing features? More specifically, under the Naive Bayes assumption, to what class does $\mathbf{x}_{\mathbf{missing}} = [1\ ?\ 1\ ?\ 1]^T$ belong to?

Same scheiße, do the calculations for the values we have

$$\begin{aligned}
p\left(C=a \mid X_1=1, X_3=1, X_5=1\right) &= \\
\frac{p(C=a)\, p\left(X_1=1, X_3=1, X_5=1 \mid C=a\right)}{p\left(X_1=1, X_3=1, X_5=1\right)} &= \\
\frac{p(C=a)\, p(X_1=1 \mid C=a)\, p(X_3=1 \mid C=a)\, p(X_5=1 \mid C=a)}{p\left(X_1=1, X_3=1, X_5=1\right)} &= \\
\frac{\frac{3}{6}\frac{3}{3}\frac{0}{3}\frac{2}{3}}{p\left(X_1=1, X_3=1, X_5=1\right)} &= \\
0 &
\end{aligned}$$

...

Again, "b" will be the only class with non-zero probability.