Lab 8
Probabilistic Reasoning - Gaussian distribution and Bayes Theorem

FTP MachLe MSE
HS 2023

Machine Learning
WÜRC

The central *paradigm of probabilistic reasoning* is to identify all relevant variables $x_1, \ldots, x_N$ in the environment,and make a probabilistic model $p(x_1, \ldots, x_N)$ of their interaction. Reasoning (*inference*) is then performed by introducing *evidence* that sets variables in known states, and subsequently computing probabilities of interest, *conditioned on this evidence*. The rules of probability, combined with Bayes' rule make a reasoning system complete.

After this unit, ...

Lernziele/Kompetenzen

- you have repeated the *basic rules of probability theory*.
- you know the difference between a *joint* and a *conditional probability* distribution.
- you know how to apply *Bayes Theorem* to calculate the *posterior* probability distribution for simple discrete examples. You can name the *prior* probability distribution, the *likelihood function*, the *evidence*, and you know how to *marginalize* over a joint probability distribution.
- you know the basic properties of a *multivariate Gaussian* probability distribution. You can plot a 2D Gaussian probability distribution given the *mean vector* $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$.
- you can *sample* data points from a given multivariate gaussian distribution.
- you can explain the *naïve Bayes classifier* to your classmates and to your teacher.

1. Supervised Bayesian Learining [M,II]

The table below contains the result of a market survey for a promotion for different items such as a magazine, a watch and a life insurance and a credit card insurance. 10 people were interviewed and asked whether they would buy such items.

Use this count table for supervised Bayesian learning. The output attribute is *sex* with possible values `male` and `female`. Consider an individual who has said *no* to the life insurance promotion, *yes* to the magazine promotion, *yes* to the watch promotion and *yes* to the credit card insurance. Use the values in the table together with the Naive Bayes classifier to determine which of a,b,c or d represents the probability that this individual is male. $p(E)$ is the marginal distribution.

	Magazine		Watch		Life Insurance		Credit Card	
	`male`	`female`	`male`	`female`	`male`	`female`	`male`	`female`
yes	4	3	2	2	2	3	2	1
no	2	1	4	2	4	1	4	3

Welche der folgenden Aussagen sind wahr und welche falsch?	**wahr**	**falsch**
a) $p(\texttt{sex} = \texttt{male}\|...) = \frac{4}{6} \cdot \frac{2}{6} \cdot \frac{2}{6} \cdot \frac{2}{6} \cdot \frac{6}{10} \cdot \frac{1}{p(E)}$	○	○
b) $p(\texttt{sex} = \texttt{male}\|...) = \frac{4}{6} \cdot \frac{2}{6} \cdot \frac{3}{4} \cdot \frac{2}{6} \cdot \frac{3}{4} \cdot \frac{1}{p(E)}$	○	○
c) $p(\texttt{sex} = \texttt{male}\|...) = \frac{4}{6} \cdot \frac{4}{6} \cdot \frac{2}{6} \cdot \frac{2}{6} \cdot \frac{6}{10} \cdot \frac{1}{p(E)}$	○	○
d) $p(\texttt{sex} = \texttt{male}\|...) = \frac{2}{6} \cdot \frac{4}{6} \cdot \frac{4}{6} \cdot \frac{2}{6} \cdot \frac{4}{10} \cdot \frac{1}{p(E)}$	○	○

2. Hamburger and Bayes Rule [A,II]

Consider the following fictitious scientic information: Doctors find that people with Kreuzfeld-Jacob disease (KJ) almost invariably ate hamburgers (Hamburger Eater, HE) , thus $p(\mathsf{HE}|\mathsf{KJ}) = 0.9$. The probability of an individual having KJ is currently rather low, about one in $100'000$.

a) Assuming eating lots of hamburgers is rather widespread, say $p(\mathsf{HE}) = 0.5$, what is the probability that a hamburger eater will have Kreuzfeld-Jacob disease? Determine the prior, the likelihood function and the posterior probability.

b) If the fraction of people eating hamburgers was rather small, $p(\mathsf{HE}) = 0,001$, what is the probability that a regular hamburger eater will have Kreuzfeld-Jacob disease?

3. Naïve Bayes Classifier [A,II]

In order to reduce my email load, I decide to implement a machine learning algorithm to decide whether or not I should read an email, or simply file it away instead. To train my model, I obtain the following data set of binary-valued *features* about each email, including whether I know the author or not, whether the email is long or short, and whether it has any of several key words, along with my final decision about whether to read it ($y = +1$ for 'read', $y = -1$

for 'discard').

x_1	x_2	x_3	x_4	x_5	y
know `author`?	is `long`?	has `research`	has `grade`	has `lottery`	$\Longrightarrow$ read?
0	0	1	1	0	-1
1	1	0	1	0	-1
0	1	1	1	1	-1
1	1	1	1	0	-1
0	1	0	0	0	-1
1	0	1	1	1	+1
0	0	1	0	0	+1
1	0	0	0	0	+1
1	0	1	1	0	+1
1	1	1	1	1	-1

a) Compute all the probabilities necessary for a naïve Bayes classifier, i.e. the class probability $p(y)$ and all the individual feature probabilities $p(x_i|y)$, for each class y and feature x_i.

b) Which class would be predicted for $x = \{00000\}$?
What about for $x = \{11010\}$?

c) Compute the *posterior probability* that $y = +1$ given the observation $x = \{00000\}$. Also compute the posterior probability that $y = +1$ given the observation $x = \{11010\}$.

d) Why should we probably not use a *'joint' Bayes classifier* (using the joint probability of the features x, as opposed to the conditional independencies assumed by naïve Bayes) for these data?

e) Suppose that before we make our predictions, we lose access to my address book, so that we cannot tell whether the email author is known. Do we need to re-train the model to classify based solely on the other four features? If so, how? *Hint:* How do the parameters of a naïve Bayes model over only features $x_2, \ldots, x_5$ differ?

4. Passenger Scanner [A, II]

A secret government agency has developed a scanner which determines whether a person is a terrorist. The scanner is fairly reliable: 95% of all scanned terrorists are identied as terrorists, and 95% of all upstanding citizens are identied as such (i.e. as non-terrorists). An informant tells the agency that exactly one passenger of 100 aboard an aeroplane in which you are seated is a terrorist. The police haul off the plane the first person for which the scanner tests positive. What is the probability that this person is a terrorist?

5. Bivariate Gaussian Distribution [A, II]

The probability density function (pdf) of a multivariate normal with $\mathbf{x} = \begin{pmatrix} x_a \\ x_b \end{pmatrix}$ and $\boldsymbol{\mu} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}$ is given by:

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}\sqrt{\det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

Consider a bivariate normal distribution $p(\mathbf{x}) = p(x_a, x_b)$ with $\mu_a = 0$, $\mu_b = 2$, $\Sigma_{aa} = 2$, $\Sigma_{bb} = 1$ and $\Sigma_{ab} = \Sigma_{ba} = \frac{\sqrt{2}}{2}$.

a) Calculate the *precision matrix* $\boldsymbol{\Lambda}$, the *inverse* $\boldsymbol{\Sigma}^{-1}$ of the covariance matrix $\boldsymbol{\Sigma}$.

b) Write out the squared generalized distance expression, the *Mahalanobis distance*

$$\boldsymbol{\Delta} = (\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}) \tag{1}$$

as a function of x_a and x_b.

c) Write out the *bivariate normal density* $p(\mathbf{x}) = p(x_a, x_b)$ (joint probability density).

d) Calculate the *eigenvalues* $\lambda_{1,2}$ and the *eigenvectors* $\mathbf{u}_{1,2}$ of the covariance matrix $\boldsymbol{\Sigma}$ using Python (`numpy.linalg.eig`).

e) Plot the joint probability distribution $p(x_a, x_b)$ using Python (`scipy.stats.multivariate_normal`) and sample $N = 10'000$ points from this distribution.

f) Calculate the *conditional probability* $p(x_a|x_b)$.

6. Weather in London [A, II]

The weather in London can be summarised as: if it rains one day there's a 70% chance it will rain the following day; if it's sunny one day there's a 40% chance it will be sunny the following day.

$$p(\textsf{today} = \textit{rain} \mid \textsf{yesterday} = \textit{rain}) = 70\%$$

$$p(\textsf{today} = \textit{sun} \mid \textsf{yesterday} = \textit{sun}) = 40\%$$

From these likelihoods, we can *infer* the following:

$$p(\textsf{today} = \textit{sun} \mid \textsf{yesterday} = \textit{rain}) = 30\%$$

$$p(\textsf{today} = \textit{rain} \mid \textsf{yesterday} = \textit{sun}) = 60\%$$

a) Assuming that the prior probability it rained yesterday is 0.5, what is the probability that it was raining yesterday given that it's sunny today?

b) If the weather follows the same pattern as above, day after day, what is the probability that it will rain on any day (based on an effectively infinte number of days of observing the weather)?

c) Use the result from b) above as a new prior probability of rain yesterday and recompute the proba- bility that it was raining yesterday given that it's sunny today.

7. Inspector Clouseau [A, II]

Inspector Clouseau arrives at the scene of a crime. The victim lies dead in the room alongside the *possible* murder weapon, a knife. The Butler (B) and Maid (M) are the inspector's main suspects and the inspector has a prior belief of 0.6 that the Butler is the murderer, and a prior belief of 0.2 that the Maid is the murderer. These beliefs are independent in the sense that $p(B, M) = p(B)p(M)$. (It is possible that both the Butler and the Maid murdered the victim or neither). The inspector's prior criminal knowledge can be formulated mathematically as follows:

$$\mathrm{dom}(B) = \mathrm{dom}(M) = \{\mathsf{murderer}, \mathsf{not\ murderer}\}$$

$$\mathrm{dom}(K) = \{\mathsf{knife\ used}, \mathsf{knife\ not\ used}\}$$

$$p(B = \mathsf{murderer}) = 0.6, \qquad p(M = \mathsf{murderer}) = 0.2$$

$$\begin{aligned}
p(\mathsf{knife\ used}|B = \mathsf{not\ murderer},\ & M = \mathsf{not\ murderer}) &= 0.3\\
p(\mathsf{knife\ used}|B = \mathsf{not\ murderer},\ & M = \mathsf{murderer}) &= 0.2\\
p(\mathsf{knife\ used}|B = \mathsf{murderer},\ & M = \mathsf{not\ murderer}) &= 0.6\\
p(\mathsf{knife\ used}|B = \mathsf{murderer},\ & M = \mathsf{murderer}) &= 0.1
\end{aligned}$$

In addition $p(K, B, M) = p(K|B, M) \cdot p(B) \cdot p(M)$.

a) Assuming that the knife is the murder weapon, what is the probability that the Butler is the murderer, i.e. what is $p(B|K)$?
Hint: Remember that it might be that neither is the murderer. Use Bayes rule and marginalize (sum) over the possible states of B and M with $\mathrm{dom}(B) = \mathrm{dom}(M) = \{\mathsf{murderer}, \mathsf{not\ murderer}\}$).

8. Factorization of a multivariate probability distribution [A, II]

By using the definition of conditional probability, show that any multivariate joint distribution of N random variables has the following trivial factorization:

$$p(x_1, x_2, \ldots, x_N) = p(x_1 \mid x_2, \ldots, x_N) \cdot p(x_2 \mid x_3, \ldots, x_N) \cdot \cdots \cdot p(x_N) \tag{2}$$

9. Conditional distribution of a bivariate Gaussian distribution [A, II]

The bivariate normal distribution is given by:

$$\mathcal{N}(\mathbf{x} \mid \mu, \mathbf{\Sigma}) = \frac{1}{\sqrt{(2\pi)^N \cdot |\mathbf{\Sigma}|}} \cdot \exp\left[-\frac{1}{2}(\mathbf{x} - \mu)^{\mathbf{T}}\mathbf{\Sigma}^{\mathbf{-1}}(\mathbf{x} - \mu)\right] \tag{3}$$

where:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{4}$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \tag{5}$$

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{6}$$

The parameter ρ is called *correlation coefficient*. By using the definition of conditional probability, show that the *conditional* distribution $p(x_1 \mid x_2)$ can be written as a *normal* distribution of the form $N(x_1 \mid \tilde{\mu}, \tilde{\sigma})$ where

$$\tilde{\mu} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2} \cdot (x_2 - \mu_2) \tag{7}$$

$$\tilde{\sigma} = (1 - \rho^2) \cdot \sigma_1^2 \tag{8}$$

Hint: Only look first at the quadratic form in the exponential of the joint probability distribution $p(x_1, x_2)$ and make use of the fact that x_2 is constant for the conditional probability distribution $p(x_1 \mid x_2)$. Sort out all terms with x_1.