CDS: Machine Learning 2023 // Tutorial Week 6

Jochem Pannekoek, Daria Mihaila & Jasper Pieterse

October 18, 2023

Graphical Models

a)

When y is observed, we can infer the probability that it is caused by disease x_i by computing the probability $p(x_i|y), i = 1, ..., n$. Exercise: give an expression for $p(x_i|y)$.

Answer:

We are only interested in the probability of a specific disease i given the observed symptom y. Since the causes are marginally independent, we can marginalize out the nuisance variables.

$$p(x_i|y) = \sum_{x_n} \frac{p(x_i, x_n, y)}{p(y)}$$

b)

Consider now the case that we have independent information that in fact the patient has disease 1 ($x_1 = 1$). Exercise: give an expression for $p(x_i|y, x_1 = 1)$ for $i \neq 1$.

Answer:

We want to find the conditional probability of the disease x_i given the symptom y and the fact that the patient has disease x_1.

$$p(x_i|y, x_1 = 1) = \frac{p(x_i, y|x_1 = 1)}{p(y|x_1 = 1)} = \frac{p(x_i, y|x_1 = 1)}{\sum_{x_i} p(x_i, y|x_1 = 1)},\ i = 2, ..., n$$

c)

We first compute the joint distribution $p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2)$ which is a Gaussian distribution because it is a product of Gaussian distributions. Therefore, it is fully specified by its mean and covariance matrix. Hint: Note, that we can express the conditional distribution of y given x_1, x_2 equivalently by saying that

$$y = \gamma(x_1 + x_2) + \xi$$

with ξ a Gaussian random variable with mean zero and variance $E\xi^2 = \sigma^2$

Answer

Given that x_1, x_2 and ξ are random Gaussian variables, the joint Gaussian distribution $p(x_1, x_2, y)$ is given by the multinomial Gaussian with a mean vector and covariance matrix we will have to derive:

$$p(x_1, x_2, y) = p(x_1)p(x_2)p(x_1, x_2|y) = \mathcal{N}(\vec{\mu}, \hat{\Sigma})$$

Mean

The mean of for y is:

$$E[y] = E[\gamma(x_1 + x_2) + \xi] = \gamma(E[x_1] + E[x_2]) + E[\xi] = \gamma(\mu_{x_1} + \mu_{x_2})$$

So the mean vector is:

$$\vec{\mu} = \begin{bmatrix} E[x_1] \\ E[x_2] \\ E[y] \end{bmatrix} = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \\ \gamma(\mu_{x_1} + \mu_{x_2}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Variance

The covariance is given by:

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

using this, we derive the components of the covariance matrix:

$Cov(x_1, x_2) = 0$ because x_1 and x_2 are independent.

$Cov(x_1, y) = Cov(y, x_1)$ can be expanded as:

$$Cov(x_1, \gamma(x_1 + x_2) + \xi) = E[x_1(\gamma(x_1 + x_2) + \xi)] - E[x_1]E[\gamma(x_1 + x_2) + \xi] \tag{1}$$

These components are evaluated as follows:

$$E[x_1(\gamma(x_1 + x_2) + \xi)] = E[\gamma x_1^2] + E[\gamma x_1 x_2] + E[x_1 \xi]$$

Given that x_1 and x_2 are independent and ξ has zero mean, we get:

$$E[x_1 \xi] = E[x_1]E[\xi] = 0$$

$$E[\gamma x_1 x_2] = E[\gamma]E[x_1]E[x_2] = \gamma \mu_{x_1} \mu_{x_2}$$

Substituting these into the equation, we get:

$$E[\gamma x_1^2 + \gamma x_1 x_2 + x_1 \xi] = \gamma E[x_1^2] + \gamma \mu_{x_1} \mu_{x_2}$$

The second term in (1) can be expanded as:

$$E[\gamma(x_1 + x_2) + \xi] = \gamma E[x_1 + x_2] + E[\xi] = \gamma(\mu_{x_1} + \mu_{x_2})$$

such that $E[x_1]E[y]$ is:

$$E[x_1]E[y] = \mu_{x_1} \gamma(\mu_{x_1} + \mu_{x_2})$$

Plugging all this into (1):

$$\begin{aligned} Cov(x_1, y) &= \gamma E[x_1^2] + \gamma \mu_{x_1} \mu_{x_2} - \mu_{x_1} \gamma(\mu_{x_1} + \mu_{x_2}) \\ &= \gamma(E[x_1^2] - \mu_{x_1}^2) \\ &= \gamma Var(x_1) \end{aligned}$$

Following similar derivation, we can also obtain

$$Cov(x_2, y) = Cov(y, x_2) = \gamma Var(x_2)$$

$Var(y)$ is given by

$$\begin{aligned} Var(y) &= Var(\gamma(x_1 + x_2) + \xi) \\ &= \gamma^2 Var(x_1 + x_2) + Var(\xi) \\ &= \gamma^2 (Var(x_1) + Var(x_2)) + \sigma^2 \end{aligned}$$

Combining this we get:

$$\hat{\Sigma} = \begin{bmatrix} Var(x_1) & 0 & \gamma Var(x_1) \\ 0 & Var(x_2) & \gamma Var(x_2) \\ \gamma Var(x_1) & \gamma Var(x_2) & \gamma^2(Var(x_1) + Var(x_2)) + \sigma^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \gamma \\ 0 & 1 & \gamma \\ \gamma & \gamma & 2\gamma^2 + \sigma^2 \end{bmatrix}$$

d)

Compute the conditional distribution $p(x_1, x_2|y)$. Hint: use the general expression for the conditional distribution of a Gaussian, as given at the end of the lecture slides

Answer

We partition our parameters in two subsets x_a and x_b with $x = (x_a, x_b)^T$ and:

$$\begin{aligned} \mu &= (\mu_a, \mu_b)^T \\ \Sigma &= \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \end{aligned}$$

the conditional distribution $p(x_a|x_b)$ is given by:

$$\begin{aligned} p(x_a|x_b) &= N(x_a|\mu_{ab}, \Sigma_{ab}) \\ \mu_{ab} &= \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b) \\ \Sigma_{ab} &= \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \end{aligned}$$

In our case, we partition $x_a = [x_1, x_2]$ and $x_b = y$. Using the mean vector and covariance matrix from the previous section, we can write their mean vectors and covariance matrices as follows:

$$\mu_a = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mu_b = 0$$

$$\Sigma_{aa} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \Sigma_{bb} = 2\gamma^2 + \sigma^2, \quad \Sigma_{ab} = \Sigma_{ba}^T = \begin{bmatrix} \gamma \\ \gamma \end{bmatrix}$$

we plug this in the formulas given to us:

$$\begin{aligned} \mu_{ab} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b) &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{1}{2\gamma^2 + \sigma^2} \begin{bmatrix} \gamma \\ \gamma \end{bmatrix} (y - 0) \\ &= \frac{\gamma}{2\gamma^2 + \sigma^2} \begin{bmatrix} y \\ y \end{bmatrix} \end{aligned}$$

$$\Sigma_{ab} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{2\gamma^2 + \sigma^2} \begin{bmatrix} \gamma \\ \gamma \end{bmatrix} \begin{bmatrix} \gamma & \gamma \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{2\gamma^2 + \sigma^2} \begin{bmatrix} \gamma^2 & \gamma^2 \\ \gamma^2 & \gamma^2 \end{bmatrix} = \begin{bmatrix} 1 - \frac{\gamma^2}{2\gamma^2+\sigma^2} & -\frac{\gamma^2}{2\gamma^2+\sigma^2} \\ -\frac{\gamma^2}{2\gamma^2+\sigma^2} & 1 - \frac{\gamma^2}{2\gamma^2+\sigma^2} \end{bmatrix}$$

Using these, we can write the conditional distribution as:

$$p(x_1, x_2|y) = N(x_a|\mu_{ab}, \Sigma_{ab})$$

e)

With Σ the covariance matrix between x_1, x_2 in the posterior distribution $p(x_1, x_2|y)$, show that the correlation coefficient $\rho = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$ is given by

$$\rho = \frac{-\gamma^2}{\gamma^2 + \sigma^2}$$

Answer

We can take the

$$\Sigma_{x_1,x_2} = \begin{bmatrix} 1 - \frac{\gamma^2}{2\gamma^2+\sigma^2} & -\frac{\gamma^2}{2\gamma^2+\sigma^2} \\ -\frac{\gamma^2}{2\gamma^2+\sigma^2} & 1 - \frac{\gamma^2}{2\gamma^2+\sigma^2} \end{bmatrix}$$

Using, $\rho = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$, we have:

$$\rho = \frac{-\frac{\gamma^2}{2\gamma^2+\sigma^2}}{\sqrt{(1 - \frac{\gamma^2}{2\gamma^2+\sigma^2})(1 - \frac{\gamma^2}{2\gamma^2+\sigma^2})}} = \frac{-\frac{\gamma^2}{2\gamma^2+\sigma^2}}{1 - \frac{\gamma^2}{2\gamma^2+\sigma^2}} = \frac{-\frac{\gamma^2}{2\gamma^2+\sigma^2}}{\frac{2\gamma^2+\sigma^2-\gamma^2}{2\gamma^2+\sigma^2}} = \frac{-\gamma^2}{\gamma^2 + \sigma^2}$$

as desired.