Homework 1

JD Gallego Posada, *University of Amsterdam* 14/09/2016

1 Probability Theory

Question 1.1

1. Define the random variables and the values they can take on, both with symbols and numerically

$$C\text{: City where I am located} \qquad C \in \mathscr{C} = \{\text{Amsterdam, Rotterdam}\}$$
$$W\text{: Weather in the city} \qquad W \in \mathscr{W} = \{\text{Rainy, Dry}\}$$

$$\mathbb{P}(W = \text{Rainy} \,|\, C = \text{Amsterdam}) = 0.5 \qquad \mathbb{P}(C = \text{Amsterdam}) = 0.8$$
$$\mathbb{P}(W = \text{Rainy} \,|\, C = \text{Rotterdam}) = 0.75 \qquad \mathbb{P}(C = \text{Rotterdam}) = 0.2$$

2. What is the probability that it does not rain when you are in Rotterdam?

$$\mathbb{P}(W = \text{Dry} \,|\, C = \text{Rotterdam}) = 1 - \mathbb{P}(W = \text{Rainy} \,|\, C = \text{Rotterdam}) = 1 - 0.75 = 0.25$$

3. What is the probability that it rains where you are?

$$\begin{aligned}\mathbb{P}(W = \text{Rainy}) &= \mathbb{P}(W = \text{Rainy} \,|\, C = \text{Rotterdam})\,\mathbb{P}(C = \text{Rotterdam}) \\ &\quad + \mathbb{P}(W = \text{Rainy} \,|\, C = \text{Amsterdam})\,\mathbb{P}(C = \text{Amsterdam}) \\ &= 0.75 \cdot 0.2 + 0.5 \cdot 0.8 = 0.55\end{aligned}$$

4. [...] What is the probability that you are in Amsterdam?

$$\begin{aligned}\mathbb{P}(C = \text{Amsterdam} \,|\, W = \text{Rainy}) &= \frac{\mathbb{P}(W = \text{Rainy} \,|\, C = \text{Amsterdam})\,\mathbb{P}(C = \text{Amsterdam})}{\mathbb{P}(W = \text{Rainy})} \\ &= \frac{0.5 \cdot 0.8}{0.55} = 0.\overline{72}\end{aligned}$$

Question 1.2

1. What is p(cancer) and p(not cancer)?

$$p(\text{cancer}) = \frac{500}{500000} = 0.001 \qquad p(\text{not cancer}) = 1 - 0.001 = 0.999$$

2. If a patient takes the blood test and it returns positive, what is the probability the patient has cancer?

$$p(\text{positive} \mid \text{cancer}) = 0.99 \qquad p(\text{positive} \mid \text{not cancer}) = 0.05$$

$$\begin{aligned}p(\text{positive}) &= p(\text{positive} \mid \text{cancer})\,p(\text{cancer}) + p(\text{positive} \mid \text{not cancer})\,p(\text{not cancer}) \\ &= 0.99 \cdot 0.001 + 0.05 \cdot 0.999 = 0.05094\end{aligned}$$

$$p(\text{cancer} \mid \text{positive}) = \frac{p(\text{positive} \mid \text{cancer})\,p(\text{cancer})}{p(\text{positive})} = \frac{0.99 \cdot 0.001}{0.05094} = 0.0194$$

3. What are some of the assumptions we are implicitly making when answering this question?

There is no error in the classification used to estimate the a priori distribution of cancer. Besides, the trials of the diagnose test were performed on a representative sample of the population, i.e., the sample was not biased towards either the group with or without cancer, for example due to reasons like age or medical pre-conditions.

Question 1.3

1. Write down the general expression for a posterior distribution, using θ for the parameter, $\mathscr{D}$ for the data. Indicate the prior, likelihood, evidence, and posterior.

$$p(\theta|\mathscr{D}) = \frac{p(\mathscr{D}|\theta)p(\theta)}{p(\mathscr{D})}$$

2. Write the posterior for this particular example. You do not need an analytic solution.

$$p(\mu\,|\,\{x_1, ..., x_N\}, \sigma^2) = \frac{\left[\prod_{i=1}^{N} \mathcal{N}(x_i\,|\,\mu, \sigma^2)\right] \mathcal{N}(\mu\,|\,\mu_0, {\sigma_0}^2)}{\int_{\mathbb{R}} \left[\prod_{i=1}^{N} \mathcal{N}(x_i\,|\,\mu', \sigma^2)\right] \mathcal{N}(\mu'\,|\,\mu_0, {\sigma_0}^2) d\mu'}$$

2 Basic Linear Algebra and Derivatives

Question 2.1

1. Compute $\boldsymbol{Ab}$

$$\mathbf{Ab} = \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 9 \\ 5 \end{bmatrix} = \begin{bmatrix} 27+25 \\ 18+15 \end{bmatrix} = \begin{bmatrix} 52 \\ 33 \end{bmatrix}$$

2. Compute $\boldsymbol{b}^T \boldsymbol{A}$

$$\mathbf{b}^T\mathbf{A} = \begin{bmatrix} 9 & 5 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 27+10 & 45+15 \end{bmatrix} = \begin{bmatrix} 37 & 60 \end{bmatrix}$$

3. What is the vector $\boldsymbol{c}$ for which $\boldsymbol{Ac} = \boldsymbol{b}$?

$$\mathbf{c} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} \text{ since } \mathbf{Ac} = \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \begin{bmatrix} -6+15 \\ -4+9 \end{bmatrix} = \begin{bmatrix} 9 \\ 5 \end{bmatrix} = \mathbf{b}$$

4. What is $\boldsymbol{A}^{-1}$?

$$\mathbf{A}^{-1} = \frac{1}{3 \cdot 3 - 2 \cdot 5} \begin{bmatrix} 3 & -5 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} -3 & 5 \\ 2 & -3 \end{bmatrix}$$

5. Verify that $\boldsymbol{A}^{-1}\boldsymbol{b} = \boldsymbol{c}$. Show that this must be the case.

$$\mathbf{A}^{-1}\mathbf{b} = \begin{bmatrix} -3 & 5 \\ 2 & -3 \end{bmatrix} \begin{bmatrix} 9 \\ 5 \end{bmatrix} = \begin{bmatrix} -27+25 \\ 18-15 \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \mathbf{c}$$

Since $\mathbf{Ac} = \mathbf{b}$ and $\mathbf{A}$ is invertible, we have $\mathbf{A}^{-1}\mathbf{Ac} = \mathbf{A}^{-1}\mathbf{b} \to \mathbf{c} = \mathbf{A}^{-1}\mathbf{b}$, as required.

Question 2.2

Find the gradient of the following functions

1. $f(x) = x^2 + 2x + 3 \to \nabla f = 2x + 2$
2. $f(x) = (2x^3 + 1)^2 \to \nabla f = 2(2x^3 + 1) \cdot 3 \cdot 2 \cdot x^2 = 12x^2(2x^3 + 1)$

Find the partial derivative of the following functions with respect to x, y, z

1. $f(x, y, z) = (x + 2y)^2 \sin(xy)$

$$\frac{\partial f}{\partial x} = 2(x + 2y)\sin(xy) + (x + 2y)^2 y\cos(xy) = (x + 2y)(2\sin(xy) + y(x + 2y)\cos(xy))$$
$$\frac{\partial f}{\partial y} = 4(x + 2y)\sin(xy) + (x + 2y)^2 x\cos(xy) = (x + 2y)(4\sin(xy) + x(x + 2y)\cos(xy))$$
$$\frac{\partial f}{\partial z} = 0$$

2. $f(x, y, z) = 2\log(x + y^2 - z)$

$$\frac{\partial f}{\partial x} = \frac{2}{x + y^2 - z} \qquad \frac{\partial f}{\partial y} = \frac{4y}{x + y^2 - z} \qquad \frac{\partial f}{\partial z} = \frac{-2}{x + y^2 - z}$$

3. $f(x, y, z) = \exp(x(\cos(y + z))$

$$\frac{\partial f}{\partial x} = \cos(y + z)\exp(x(\cos(y + z)) \qquad \frac{\partial f}{\partial y} = -x\sin(x + y)\exp(x(\cos(y + z)) = \frac{\partial f}{\partial z}$$

Question 2.3

1. *Expand the expression and gather terms.*

$$\alpha = (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \boldsymbol{S}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) =$$
$$(\boldsymbol{x}^T - \boldsymbol{\mu}^T) \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{\mu}^T - \boldsymbol{\mu}_0^T) \boldsymbol{S}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) =$$
$$(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}) (\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{\mu}^T\boldsymbol{S}^{-1} - \boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}) (\boldsymbol{\mu} - \boldsymbol{\mu}_0) =$$
$$\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{S}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^T\boldsymbol{S}^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}\boldsymbol{\mu}_0 =$$
$$\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}^T\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{S}^{-1}\boldsymbol{\mu}_0\right) - \left(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1} + \boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}\right)\boldsymbol{\mu} + \boldsymbol{\mu}^T\left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right)\boldsymbol{\mu}$$

2. *Collect all the terms that depend on $\boldsymbol{\mu}$ and those that do not.*

$$\alpha = \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}^T\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{S}^{-1}\boldsymbol{\mu}_0\right) - \left(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1} + \boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}\right)\boldsymbol{\mu} + \boldsymbol{\mu}^T\left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right)\boldsymbol{\mu}$$

3. *Take the derivative with respect to $\boldsymbol{\mu}$, set to 0, and solve for $\boldsymbol{\mu}$.*

$$\begin{aligned}
\frac{\partial \alpha}{\partial \boldsymbol{\mu}} &= -\frac{\partial}{\partial \boldsymbol{\mu}}\boldsymbol{\mu}^T\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{S}^{-1}\boldsymbol{\mu}_0\right) - \frac{\partial}{\partial \boldsymbol{\mu}}\left(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1} + \boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}\right)\boldsymbol{\mu} + \frac{\partial}{\partial \boldsymbol{\mu}}\boldsymbol{\mu}^T\left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right)\boldsymbol{\mu} \\
&= -\left(\boldsymbol{x}^T(\boldsymbol{\Sigma}^{-1})^T + \boldsymbol{\mu}_0^T(\boldsymbol{S}^{-1})^T\right) - \left(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1} + \boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}\right) + \boldsymbol{\mu}^T\left[\left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right) + \left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right)^T\right] \\
&= -2\left(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1} + \boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}\right) + 2\boldsymbol{\mu}^T\left(\boldsymbol{\Sigma}^{-1} + \boldsymbol{S}^{-1}\right) = 0
\end{aligned}$$

$$2\boldsymbol{\mu}^T\left(\boldsymbol{\Sigma}^{-1}+\boldsymbol{S}^{-1}\right)=2\left(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}+\boldsymbol{\mu}_0^T\boldsymbol{S}^{-1}\right)$$
$$\left(\boldsymbol{\Sigma}^{-1}+\boldsymbol{S}^{-1}\right)\boldsymbol{\mu}=\boldsymbol{\Sigma}^{-1}\boldsymbol{x}+\boldsymbol{S}^{-1}\boldsymbol{\mu}_0$$

Assuming $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$ are positive definite matrices, $det(\boldsymbol{\Sigma})$, $det(\boldsymbol{S})$, $det(\boldsymbol{\Sigma}^{-1})$, and $det(\boldsymbol{S}^{-1})$ are all strictly greater than 0. Then, by the Minkowski determinant theorem,

$$det(\boldsymbol{\Sigma}^{-1}+\boldsymbol{S}^{-1})\geq det(\boldsymbol{\Sigma}^{-1})+det(\boldsymbol{S}^{-1})>0.$$

Thus, $(\boldsymbol{\Sigma}^{-1}+\boldsymbol{S}^{-1})$ is also invertible and it is possible to solve for $\boldsymbol{\mu}$:

$$\boldsymbol{\mu}=\left(\boldsymbol{\Sigma}^{-1}+\boldsymbol{S}^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{x}+\boldsymbol{S}^{-1}\boldsymbol{\mu}_0\right)$$