Machine Learning- 2019-2029
http://math.unice.fr/~rubentha/cours.html

Exam for MathMods, MPA→ANSWERS

Exercise 1.

(1) We have

$$\mathbf{X}^T = [\mathbf{x}_1, \dots, \mathbf{x_N}]$$

(it is a $2 \times N$ matrix). So

$$\begin{aligned} \mathbf{X}^T\mathbf{t} &= \begin{bmatrix} x_{1,1}t_1 + x_{2,1}t_2 + \dots + x_{N,1}t_N \\ x_{1,2}t_1 + x_{2,2}t_2 + \dots + x_{N,2}t_N \end{bmatrix} \\ &= \sum_{n=1}^{N} \mathbf{x}_n t_n \,. \end{aligned}$$

(2) We have

$$\begin{aligned} \mathbf{X}^T\mathbf{X}\mathbf{w} &= \begin{bmatrix} \sum_{n=1}^N x_{n,1}^2 & \sum_{n=1}^N x_{n,1}x_{n,2} \\ \sum_{n=1}^N x_{n,1}x_{n,2} & \sum_{n=1}^N x_{n,2}^2 \end{bmatrix} \times [w_0, w_1] \\ &= \begin{bmatrix} \sum_{n=1}^N x_{n,1}^2 & \sum_{n=1}^N x_{n,1}x_{n,2} \\ \sum_{n=1}^N x_{n,1}x_{n,2} & \sum_{n=1}^N x_{n,2}^2 \end{bmatrix} \\ &= \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^T \mathbf{w} \,. \end{aligned}$$

Exercise 2.

(1) We have

$$\begin{aligned} p(\mu|x_1,\dots,x_N) &\propto p(x_1,\dots,x_N|\mu)p(\mu) \\ &= \prod_{i=1}^{N} \frac{e^{-\frac{1}{2}(x_i-\mu)^2}}{\sqrt{2\pi}} \times \frac{e^{-\frac{1}{2A}(\mu-\mu_0)^2}}{\sqrt{2\pi A}} \\ &\propto \exp\Bigg[-\frac{1}{2}\left(N+\frac{1}{A}\right)\left(\mu - \frac{\mu_0}{A}(N+1/A)^{-1} - (N+1/A)^{-1}\sum_{i=1}^{N} x_i\right)^2 \\ &\quad +\frac{1}{2}\left(N+\frac{1}{A}\right)^{-1}\frac{\mu_0^2}{A^2} + \frac{1}{2}\left(N+\frac{1}{A}\right)^{-1}\left(\sum_{i=1}^{N} x_i\right)^2\Bigg] \\ &\propto \frac{1}{\sqrt{2\pi}\sqrt{(N+1/A)^{-1}}} \exp\left[-\frac{1}{2}\left(N+\frac{1}{A}\right)\left(\mu - \frac{\mu_0}{A}(N+1/A)^{-1} - (N+1/A)^{-1}\sum_{i=1}^{N} x_i\right)^2\right] . \end{aligned}$$

So the desired law is

$$\mathcal{N}\left(\frac{\mu_0}{A}(N+1/A)^{-1} + (N+1/A)^{-1}\sum_{i=1}^{N} x_i, (N+1/A)^{-1}\right) .$$

(2) The computation in dimension d is similar. We have (with $\overline{x} = \sum_{i=1}^N x_i$)

$$p(\mu|x_1,\dots,x_N) = p(x_1,\dots,x_N|\mu)p(\mu)$$
$$= \prod_{i=1}^N \frac{e^{-\frac{1}{2}(x_i-\mu)^T(x_i-\mu)}}{(2\pi)^{d/2}} \times \frac{e^{-\frac{1}{2}(\mu-\mu_0)^T A^{-1}(\mu-\mu_0)}}{(2\pi)^{d/2}\sqrt{|\det(A)|}}$$
$$\propto \exp\left(-\frac{1}{2}\left(\mu - \left(NI_d + A^{-1}\right)^{-1}\left(\overline{x} + A^{-1}\mu_0\right)\right)^T \left(NI_d + A^{-1}\right)\left(\mu - \left(NI_d + A^{-1}\right)^{-1}\left(\overline{x} + A^{-1}\mu_0\right)\right)\right).$$

So the desired law is

$$\mathcal{N}\left(\left(NI_d + A^{-1}\right)^{-1}\left(\overline{x} + A^{-1}\mu_0\right), \left(NI_d + A^{-1}\right)^{-1}\right).$$

Exercise 3. We want to find the minimum of Φ under the constraint : $\Psi(\mathbf{z}) := \sum_{i=1}^n z_i = 1$. We have

$$\nabla\Phi(\mathbf{z}) = \begin{bmatrix} -y_1/z_1 \\ \vdots \\ -y_n/z_n \end{bmatrix},$$

and

$$\nabla\Psi(\mathbf{z}) = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

We look for a $\mathbf{z}$ in S such that $\nabla\Phi(\mathbf{z})$ and $\nabla\Psi(\mathbf{z})$ are colinear. That is :

$$\exists\lambda \text{ such that } y_i = \lambda z_i \,,\ \forall i\,,$$
$$\text{and } \sum_{i=1}^n y_i = \sum_{i=1}^n z_i = 1\,.$$

This leads to $y_i = z_i$ ($\forall i$).

Let us now show that Φ is convex. For $\mathbf{z}$, $\mathbf{z}'$ in S and $\lambda \in [0,1]$, we have

$$\begin{aligned} \Phi(\lambda\mathbf{z} + (1-\lambda)\mathbf{z}') &= \sum_{i=1}^n -y_i \log(\lambda z_i + (1-\lambda)z_i') \\ \text{(because log is concave)} &\leq \sum_{i=1}^n -y_i(\lambda\log(z_i) + (1-\lambda)\log(z_i')) \\ &= \lambda\Phi(\mathbf{z}) + (1-\lambda)\Phi(\mathbf{z}')\,. \end{aligned}$$

As Φ is convex, the critical point we have found is an absolute minimum.

Exercise 4.

(1) We have

$$p(x_1,\dots,x_N|\mu,\Sigma) = \prod_{i=1}^N \frac{e^{-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)}}{(2\pi)^{3/2}(\det(\Sigma))^{1/2}}\,.$$

(2) When $x_1,\dots,x_N$ is fixed, we set $\Phi(\mu,\Sigma) = p(x_1,\dots,x_N|\mu,\Sigma)$. We want to find the maximum of Φ. This is the same as the maximum of

$$\mathcal{L}(\mu,\Sigma) = \log(p(x_1,\dots,x_N|\mu,\Sigma)) = -\frac{N}{2}\log((2\pi)^3\det(\Sigma)) - \sum_{i=1}^N \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\,.$$

We have the following differentials (for all matrix H, for all vector h)

$$(\det)'(\Sigma).H = \mathrm{Tr}(\mathrm{Com}(\Sigma)^T H)\,,$$
$$(\Sigma + H)^{-1} = \Sigma^{-1} - \Sigma^{-1}H\Sigma^{-1} + o(H)\,,$$
$$\mathcal{L}(\mu + h, \Sigma) = \sum_{i=1}^N \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}h\,.$$

We look for a point where the differentials are zero. So μ should be

$$\mu = \overline{x} := \frac{1}{N}\sum_{i=1}^{N} x_i\,. \tag{0.1}$$

Then for all matrix H, we have

$$\mathcal{L}(\mu, \Sigma + H) = \mathcal{L}(\mu, \Sigma) - \frac{N}{2}\frac{\mathrm{Tr}(\mathrm{Com}(\Sigma)^T H)}{\det(\Sigma)} + \frac{1}{2}\sum_{k=1}^{N}(x_k - \mu)^T \Sigma^{-1} H \Sigma^{-1}(x_k - \mu) + o(H)\,.$$

We want the first order term to be zero for all H in $\mu = \overline{x}$, that is :

$$\frac{N}{2}\frac{\mathrm{Tr}(\mathrm{Com}(\Sigma)^T H)}{\det(\Sigma)} = \frac{1}{2}\sum_{k=1}^{N}(x_k - \overline{x})^T \Sigma^{-1} H \Sigma^{-1}(x_k - \overline{x})\,. \tag{0.2}$$

This implies that for all H', (remember $\mathrm{Com}(\Sigma)^T\Sigma = \det(\Sigma)$)

$$\begin{aligned}\frac{N}{2}\frac{\mathrm{Tr}(\mathrm{Com}(\Sigma)^T \Sigma H' \Sigma)}{\det(\Sigma)} &= \frac{1}{2}\sum_{k=1}^{N}(x_k - \overline{x})^T H'(x_k - \overline{x}) \\ \mathrm{Tr}(H'\Sigma) &= \frac{1}{N}\sum_{k=1}^{N}(x_k - \overline{x})^T H'(x_k - \overline{x})\,.\end{aligned}$$

With $H' = ((\delta_{j,i}))_{1\le i,j\le 3}$, we obtain (for all i, j)

$$\Sigma_{i,j} = \frac{1}{N}\sum_{k=1}^{N}(x_k - \overline{x})_i (x_k - \overline{x})_j\,.$$

So

$$\Sigma = \Sigma_0 := \frac{1}{N}\sum_{k=1}^{N}(x_k - \overline{x})(x_k - \overline{x})^T\,. \tag{0.3}$$

Under the assumption that $(x_1, \ldots, x_N)$ is a basis, this Σ is invertible (easy exercise) so Equation (0.2) is satisfied

Exam for IM, EDHEC→ANSWERS

Documents and calculators are not allowed. The grading will be function of your justifications. The exercises are independent.

Exercise 1.

(1) We have
$$\mathbf{X}^T = [\mathbf{x}_1, \dots, \mathbf{x}_\mathbf{N}]$$
(it is a $2 \times N$ matrix). So
$$\begin{aligned}\mathbf{X}^T\mathbf{t} &= \left[\begin{array}{c} x_{1,1}t_1 + x_{2,1}t_2 + \dots + x_{N,1}t_N \\ x_{1,2}t_1 + x_{2,2}t_2 + \dots + x_{N,2}t_N \end{array}\right] \\ &= \sum_{n=1}^{N} \mathbf{x}_n t_n \,.\end{aligned}$$

(2) We have
$$\begin{aligned}\mathbf{X}^T\mathbf{X}\mathbf{w} &= \left[\begin{array}{cc} \sum_{n=1}^N x_{n,1}^2 & \sum_{n=1}^N x_{n,1}x_{n,2} \\ \sum_{n=1}^N x_{n,1}x_{n,2} & \sum_{n=1}^N x_{n,2}^2 \end{array}\right] \times [w_0, w_1] \\ &= \left[\begin{array}{cc} \sum_{n=1}^N x_{n,1}^2 & \sum_{n=1}^N x_{n,1}x_{n,2} \\ \sum_{n=1}^N x_{n,1}x_{n,2} & \sum_{n=1}^N x_{n,2}^2 \end{array}\right] \\ &= \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^T \mathbf{w} \,.\end{aligned}$$

Exercise 2. We have
$$\begin{aligned}p(\mu|x_1,\dots,x_N) &\propto p(x_1,\dots,x_N|\mu)p(\mu) \\ &= \prod_{i=1}^N \frac{e^{-\frac{1}{2}(x_i-\mu)^2}}{\sqrt{2\pi}} \times \frac{e^{-\frac{1}{2A}(\mu-\mu_0)^2}}{\sqrt{2\pi A}} \\ &\propto \exp\Bigg[-\frac{1}{2}\left(N+\frac{1}{A}\right)\left(\mu - \frac{\mu_0}{A}(N+1/A)^{-1} - (N+1/A)^{-1}\sum_{i=1}^N x_i\right)^2 \\ &\quad +\frac{1}{2}\left(N+\frac{1}{A}\right)^{-1}\frac{\mu_0^2}{A^2} + \frac{1}{2}\left(N+\frac{1}{A}\right)^{-1}\left(\sum_{i=1}^N x_i\right)^2\Bigg] \\ &\propto \frac{1}{\sqrt{2\pi}\sqrt{(N+1/A)^{-1}}} \exp\left[-\frac{1}{2}\left(N+\frac{1}{A}\right)\left(\mu - \frac{\mu_0}{A}(N+1/A)^{-1} - (N+1/A)^{-1}\sum_{i=1}^N x_i\right)^2\right] .\end{aligned}$$
So the desired law is
$$\mathcal{N}\left(\frac{\mu_0}{A}(N+1/A)^{-1} + (N+1/A)^{-1}\sum_{i=1}^N x_i, (N+1/A)^{-1}\right) .$$

Exercise 3. We write the answer for $d = 1$ (this is not altogether correct).

(1) We have
$$p(x_1,\dots,x_N|\mu,\Sigma) = \prod_{i=1}^N \frac{e^{-\frac{1}{2\Sigma}(x_i-\mu)^2}}{\sqrt{2\pi\Sigma}} \,.$$

(2) When $x_1,\dots,x_N$ is fixed, we set $\Phi(\mu,\Sigma) = p(x_1,\dots,x_N|\mu,\Sigma)$. We want to find the maximum of Φ. This is the same as the maximum of
$$\mathcal{L}(\mu,\Sigma) = \log(p(x_1,\dots,x_N|\mu,\Sigma)) = -\frac{N}{2}\log(2\pi\Sigma) - \sum_{i=1}^N \frac{1}{2\Sigma}(x_i-\mu)^2 \,.$$

We compute

$$\frac{\partial \mathcal{L}}{\partial \mu}(\mu, \Sigma) = \sum_{i=1}^{N} \frac{(x_i - \mu)}{\Sigma} \,,$$

$$\frac{\partial \mathcal{L}}{\partial \Sigma}(\mu, \Sigma) = -\frac{N}{2\Sigma} + \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\Sigma^2} \,.$$

So, there is only one point where $\nabla \mathcal{L}$ is zero and that is

$$\mu_0 = \frac{1}{N} \sum_{i=1}^{N} x_i \,,\ \Sigma_0 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_0)^2 \,.$$