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

Exam for MathMods, MPA

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

Exercise 1. [6 points] When multiplying a scalar by a vector (or matrix), we multiply each element of the vector (or matrix) by that scalar. For $\mathbf{x}_n = [x_{n,1}, x_{n,2}]^T$ ($1 \leq n \leq N$), $\mathbf{t} = [t_1, \dots, t_N]^T$, $\mathbf{w} = [w_0, w_1]$ and

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix},$$

(1) show that

$$\sum_{n=1}^{N} \mathbf{x}_n t_n = \mathbf{X}^T \mathbf{t},$$

(2) and

$$\sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^T \mathbf{w} = \mathbf{X}^T \mathbf{X} \mathbf{w}.$$

Exercise 2. [6 points] We have an a priori law on a parameters μ (in $\mathbb{R}^d$) :

$$p(\mu) = \mathcal{N}(\mu_0, A),$$

where A is a $d \times d$ matrix, definite, positive. We suppose that $\Sigma = I_d$ (identity matrix of size d). We observe $x_1, \dots, x_N$ (N in $\mathbb{N}^*$), independent and identically distributed, of law $\mathcal{N}(\mu, \Sigma)$. We want to compute the a posteriori law on μ (knowing these observations). We write $p(\mu | x_1, \dots, x_N)$ for this a posteriori law.

(1) We suppose first that $d = 1$ (A is then a real number). Compute $p(\mu | x_1, \dots, x_n)$.
(2) We suppose d is in $\mathbb{N}^*$. Remember that if X is of law $\mathcal{N}(m, B)$ in $\mathbb{R}^d$, then its density is

$$x \in \mathbb{R}^d \mapsto \frac{e^{-\frac{1}{2} x^T B^{-1} x}}{(2\pi)^{d/2} \sqrt{|\det(B)|}}.$$

Compute $p(\mu | x_1, \dots, x_n)$.

Exercise 3. [6 points] For $\mathbf{z} \in (\mathbb{R}^+)^n$, $\mathbf{y} \in (\mathbb{R}^+)^n$ such that $\sum_{i=1}^n y_i = 1$, $\sum_{i=1}^n z_i = 1$, we define

$$H(y, z) = -\sum_{i=1}^{n} y_i \log(z_i).$$

We set $S = \{\mathbf{z} \in (\mathbb{R}^+)^n : \sum_{i=1}^n z_i = 1\}$. We fix $\mathbf{y}$ in S and define

$$\Phi : \mathbf{z} \in S \mapsto H(\mathbf{y}, \mathbf{z}).$$

Show that $\mathbf{z} = \mathbf{y}$ is an absolute minimum of Φ.

Exercise 4. [6 points] We observe $\mathbf{x}_1, \dots, \mathbf{x}_N$ in $\mathbb{R}^d$ ($d = 3$), variables of law $\mathcal{N}(\mu, \Sigma)$ (independent).

(1) Write the density of the law $p(\mathbf{x_1}, \dots, \mathbf{x}_N | \mu, \Sigma)$.
(2) What is the max-likelihood estimator of μ, Σ, knowing $\mathbf{x}_1, \dots, \mathbf{x}_N$? (To solve this question, you should find a critical point. After what you can assume that this point is a maximum.)

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

Exam for IM, EDHEC

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

Exercise 1. [7 points] When multiplying a scalar by a vector (or matrix), we multiply each element of the vector (or matrix) by that scalar. For $\mathbf{x}_n = [x_{n,1}, x_{n,2}]^T$ $(1 \leq n \leq N)$, $\mathbf{t} = [t_1, \dots, t_N]^T$, $\mathbf{w} = [w_0, w_1]$ and

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix},$$

(1) show that

$$\sum_{n=1}^{N} \mathbf{x}_n t_n = \mathbf{X}^T \mathbf{t},$$

(2) and

$$\sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^T \mathbf{w} = \mathbf{X}^T \mathbf{X} \mathbf{w}.$$

Exercise 2. [7 points] We have an a priori law on a parameters μ (in $\mathbb{R}$) :

$$p(\mu) = \mathcal{N}(\mu_0, A),$$

where A is in $\mathbb{R}^{+*}$. We suppose that $\Sigma = 1$ (identity matrix of size d). We observe $x_1, \dots, x_N$ (N in $\mathbb{N}^*$), independent and identically distributed, of law $\mathcal{N}(\mu, \Sigma)$. We want to compute the a posteriori law on μ (knowing these observations). We write $p(\mu|x_1, \dots, x_N)$ for this a posteriori law.

Compute $p(\mu|x_1, \dots, x_n)$.

Exercise 3. [7 points] We observe $\mathbf{x}_1, \dots, \mathbf{x}_N$ in $\mathbb{R}^d$ $(d = 3)$, variables of law $\mathcal{N}(\mu, \Sigma)$ (independent). We suppose that the system $(x_1, \dots, x_N)$ is a basis of $\mathbb{R}^d$ (this happens with probability 1).

(1) Write the density of the law $p(\mathbf{x_1}, \dots, \mathbf{x}_N|\mu, \Sigma)$.
(2) What is the max-likelihood estimator of μ, Σ, knowing $\mathbf{x}_1, \dots, \mathbf{x}_N$? (To solve this question, you should find a critical point. After what you can assume that this point is a maximum.)