Université Nice Sophia-Antipolis
Statistical learning, 2021-2022
`https://math.unice.fr/~rubentha/cours.html`

Final examination (MPA & MathMods), B

Documents and calculators forbidden. Give back the subject with your copy (+0.5 points!).
Duration: 2h.

Part 1.

(1) a
(2) c
(3) a
(4) c
(5) b
(6) c
(7) d
(8) b
(9) c
(10) a

Part 2. **Mathematics exercises (all exercises are independent)**

Exercise 1. See Section 8.5.2 of the poly.

Exercise 2. We call $X_1, \dots, X_N$ the N points. For all i, $m \in [0, 1/2]$, $\mathbb{P}(\|X_i\| \leq m) = v_p m^p$. For all m,

$$\begin{aligned}
\mathbb{P}(R > m) &= \mathbb{P}(\|X_1\| > m, \dots, \|X_N\| > m) \\
\text{(independance)} &= \prod_{i=1}^{N} \mathbb{P}(\|X_i\| > m) \\
&= \prod_{i=1}^{N} (1 - \mathbb{P}(\|X_i\| \leq m)) \\
&= (1 - v_p m^p)^N .
\end{aligned}$$

So the median of R is m_0 such that

$$\begin{aligned}
(1 - v_p m_0^p)^N &= \frac{1}{2} \\
v_p m_0^p &= 1 - \left(\frac{1}{2}\right)^{1/N} \\
m_0 &= v_p^{1/p} \left(1 - \left(\frac{1}{2}\right)^{1/N}\right)^{1/p} .
\end{aligned}$$

Exercise 3.

(1) The i-th line in Xw is $\sum_{j=1}^{D} x_j^{(i)} w_j$. So

$$\left(\begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix} - Xw\right)^T \left(\begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix} - Xw\right) = \sum_{i=1}^{N} \left(t_i - \sum_{j=1}^{D} x_j^{(i)} w_j\right)^2$$

$$= \sum_{i=1}^{N} (t_i - w^T x^{(i)})^2 \,.$$

(2) For all j in $\{1, 2, \ldots, D\}$,

$$\begin{aligned} \frac{1}{2}\frac{\partial \mathcal{L}}{\partial w_j} &= \sum_{i=1}^{N} (w^T x^{(i)} - t_i) x_j^{(i)} \\ &= \sum_{i=1}^{N} \sum_{k=1}^{D} w_k x_k^{(i)} x_j^{(i)} - \sum_{i=1}^{N} t_i x_j^{(i)} \\ &= \sum_{i=1}^{N} x_j^{(i)} (Xw)_i - (X^T \boldsymbol{t})_j \\ &= (X^T X w)_j - (X^T \boldsymbol{t})_j \,. \end{aligned}$$

So

$$\frac{1}{2} \nabla \mathcal{L}(w) = (X^T X) w - X^T \boldsymbol{t} \,.$$

(3) The gradient of $\mathcal{L}$ is zero for

$$w = (X^T X)^{-1} X^T \boldsymbol{t} \,.$$

For each i,

$$w \mapsto (w^T x^{(i)} - t_i)^2 = (\sum_{j=1}^{D} w_j x_j^{(i)} - t_i)^2$$

is convex (as a polynomial of degree two in $w_1, \ldots, w_D$, with positive coefficients for the w_j^2, no double product $w_{j_1} w_{j_2}$). The function $\mathcal{L}$ is convex as a sum of convex functions. So this point is the absolute minimum of $\mathcal{L}$.

(4) See Exercise 2.1, Section 2.3.2 of the poly.