

Answers for final examination (IM), B

Documents and calculators forbidden. Give back the subject with your copy (+0.5 points!).
 Duration: 2h30.

Part 1.

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

Part 2.

Exercise 1. See Section 8.5.2 of the poly.

Exercise 2.

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

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

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

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

So

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

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

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

For each i ,

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

is convex (as a polynomial of degree two in w_1, \dots, 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} .