

Question 1:

Determine whether each of the following scalar-valued functions of n -vectors is linear. If it is a linear function, give its inner product representation, ie., an n -vector \mathbf{a} for which $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x}$ for all \mathbf{x} . If it is not linear, give specific \mathbf{x}, \mathbf{y} , α and β such that

$$f(\alpha \mathbf{x} + \beta \mathbf{y}) \neq \alpha f(\mathbf{x}) + \beta f(\mathbf{y}).$$

- (a) The spread of values of the vector, defined as $f(\mathbf{x}) = \max_k x_k - \min_k x_k$.
(b) The difference of the last element and the first, $f(\mathbf{x}) = x_n - x_1$.

Answer :

(a)

Take $\mathbf{x} = (1, 2, 3)$ and $\alpha = 1, \beta = 1$ for example:

$$\begin{aligned} f(\mathbf{x}) &= 3 - 1 = 2 \\ f(-\mathbf{x}) &= -1 + 3 = 2 \\ f(\mathbf{0}) &= 0 - 0 = 0 \\ f(\mathbf{x} + (-\mathbf{x})) &= f(\mathbf{0}) = 0 \\ f(\mathbf{x}) + f(-\mathbf{x}) &= 2 + 2 = 4 \\ f(\mathbf{x} + (-\mathbf{x})) &\neq f(\mathbf{x}) + f(-\mathbf{x}) \end{aligned}$$

In conclusion, $f(\mathbf{x}) = \max_k x_k - \min_k x_k$ is not a linear function.

(b)

We know:

$$\alpha \mathbf{x} + \beta \mathbf{y} = (\alpha x_1 + \beta y_1, \dots, \alpha x_n + \beta y_n)$$

$$\begin{aligned} f(\mathbf{x}) &= x_n - x_1 \\ f(\mathbf{y}) &= y_n - y_1 \end{aligned}$$

$$\begin{aligned} f(\alpha \mathbf{x} + \beta \mathbf{y}) &= \alpha x_n + \beta y_n - (\alpha x_1 + \beta y_1) \\ \alpha f(\mathbf{x}) + \beta f(\mathbf{y}) &= \alpha(x_n - x_1) + \beta(y_n - y_1) \\ &= \alpha x_n + \beta y_n - (\alpha x_1 + \beta y_1) \\ f(\alpha \mathbf{x} + \beta \mathbf{y}) &= \alpha f(\mathbf{x}) + \beta f(\mathbf{y}). \end{aligned}$$

Let's denote \mathbf{e}_i as the vector in \mathbb{R}^n where the i -th entry is equal to 1, and all other entries are equal to 0.

$$f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} = (\mathbf{e}_n - \mathbf{e}_1)^T \mathbf{x}$$

In conclusion, $f(\mathbf{x}) = x_n - x_1$ is a linear function.

Question 2:

Consider the regression model $y = \mathbf{x}^T \mathbf{a} + b$, where y is the predicted response, \mathbf{x} is an 8-vector of features, \mathbf{a} is an 8-vector of coefficients, and b is the offset term. Determine with reasoning whether each of the following statements is true or false.

- (a) If $a_3 > 0$ and $x_3 > 0$, then $y \geq 0$
- (b) If $a_2 = 0$ then the prediction y does not depend on the second feature x_2 .
- (c) If $a_6 = -0.8$, then increasing x_6 (keeping all other x is the same) will decrease y .

Answer:

(a) False.

From the condition, we can deduce that $a_3 x_3 > 0$. but we can not deduce $\sum_{i=1, i \neq 3}^8 a_i x_i > 0$ and $b > 0$. Thus, we can not ensure $y = \sum_{i=1, i \neq 3}^8 a_i x_i + b + a_3 x_3 > 0$.

(b) True.

From the condition, we can deduce that $y = \sum_{i=1, i \neq 2}^8 a_i x_i + b$, which implies that y does not depend on the second feature x_2

(c) True.

Assume $x'_6 = x_6 + d, d > 0$, we know $y' = \sum_{i=0}^8 a_i x_i + d = y + d$. $y' - y = d > 0$

We can conclude that increasing x_6 will decrease y .

Question 3:

In linear regression models, we consider two data points (\mathbf{x}_1, y_1) and (\mathbf{x}_2, y_2) with $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$ and $y_1, y_2 \in \mathbb{R}$. For simplicity, we set the bias term $b = 0$. Let $\mathbf{X} \in \mathbb{R}^{2 \times 2}$ have rows \mathbf{x}_1^T and \mathbf{x}_2^T , and let $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in \mathbb{R}^2$. Assume the columns of \mathbf{X} , denoted by $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, are linearly dependent such that $\mathbf{x}^{(1)} = 2\mathbf{x}^{(2)}$.

(a) Consider the least squares estimation:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2^2 \quad (1)$$

What problem does the linear dependency among the columns of \mathbf{X} cause when estimating $\boldsymbol{\beta}$ using least squares?

(b) Now consider the ridge regression, which incorporates a regularization term:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2, \quad (2)$$

where $\lambda > 0$ is a regularization parameter. Derive the solution $\hat{\boldsymbol{\beta}}$ of (2). What is the ratio between $\hat{\beta}_1$ and $\hat{\beta}_2$?

(c) Discuss how varying the value of λ affects the solution and its ability to mitigate issues arising from linear dependency of columns of \mathbf{X} .

Answer

(a) According to Linear Algebra, it's obvious that \mathbf{X} does not have full column rank.

This leads to the solution of $\mathbf{X}\boldsymbol{\beta} - \mathbf{y}$ (i.e. $\boldsymbol{\beta}$ fullfilthisequation) is non-unique.

Question 4:

Let $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ be given with $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$. Consider the soft-SVM:

$$\min_{\mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}} \sum_{i=1}^N h(y_i(\langle \mathbf{a}, \mathbf{x}_i \rangle) + b) - 1 + \lambda \|\mathbf{a}\|_2^2,$$

where $\lambda \in \mathbb{R}$ is a regularization parameter and $h(t) = \max\{0, -t\}$ is the hinge loss function. Prove that solving the above soft-SVM is equivalent to solving the following problem:

$$\begin{aligned} \min_{\mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}, \xi \in \mathbb{R}^N} \quad & \sum_{i=1}^N \xi + \lambda \|\mathbf{a}\|_2^2, \\ \text{s.t.} \quad & y_i(\langle \mathbf{a}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i = 1, 2, \dots, N \end{aligned}$$

Answer

Question 5:

Let V be a Hilbert space. Let S_1 and S_2 be two hyperplanes in V defined by

$$S_1 = \{x \in V | \langle \mathbf{a}_1, x \rangle = b_1\}, S_2 = \{x \in V | \langle \mathbf{a}_2, x \rangle = b_2\}.$$

Assume $S_1 \cap S_2$ is non-empty. Let $\mathbf{y} \in V$ be given. We consider the projection of \mathbf{y} onto $S_1 \cap S_2$, i.e., the solution of

$$\min_{\mathbf{x} \in S_1 \cap S_2} \|\mathbf{x} - \mathbf{y}\|. \quad (3)$$

(a) Prove that $S_1 \cap S_2$ is a plane, i.e., if $\mathbf{x}, \mathbf{z} \in S_1 \cap S_2$, then $(1+t)\mathbf{z} - t\mathbf{x} \in S_1 \cap S_2$ for any $t \in \mathbb{R}$.

(b) Prove that \mathbf{z} is a solution of (3) if and only if $\mathbf{z} \in S_1 \cap S_2$ and

$$\langle \mathbf{z} - \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle = 0, \forall \mathbf{x} \in S_1 \cap S_2 \quad (4)$$

(c) Find an explicit solution of (3).

(d) Prove the solution found in part (c) is unique.