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 $\boldsymbol{a}$ for which $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x}$ for all $\boldsymbol{x}$. If it is not linear, give specific $\boldsymbol{x}, \boldsymbol{y}$, α and β such that

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

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

Answer :

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

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

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

(b)
We know:

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

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

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

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

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

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

Question 2:

Consider the regression model $y = \boldsymbol{x}^T\boldsymbol{a} + b$, where y is the predicted response, $\boldsymbol{x}$ is an 8-vector of features, $\boldsymbol{a}$ is an 8-vector of coefficients, and b is the offest 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 $a_3x_3 > 0$. but we can not deduce $\sum_{i=1,i\neq 3}^{8} a_ix_i > 0$ and $b > 0$. Thus, we can not ensure $y = \sum_{i=1,i\neq 3}^{8} a_ix_i + b + a_3b_3 > 0$.

(b) True.

From the condition, we can deduce that $y = \sum_{i=1,i\neq 2}^{8} a_ix_i + b$, 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 $y' = \sum_{i=0}^{8} a_ix_i + d = y + d$. $y' - y = d > 0$

We can conclude that increasing x_6 will decrease y.

Question 3:

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

(a) Consider the least squares estimation:

$$\min_{\beta\in\mathbb{R}^2} \|\boldsymbol{X}\beta - \boldsymbol{y}\|_2^2 \tag{1}$$

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

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

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

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

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

Answer

todo

Question 4:

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

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

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:

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

Answer

Question 5:

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

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

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

$$\min_{\boldsymbol{x} \in S_1 \cap S_2} \|\boldsymbol{x} - \boldsymbol{y}\|. \tag{3}$$

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

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

$$\langle \boldsymbol{z} - \boldsymbol{y}, \boldsymbol{z} - \boldsymbol{x} \rangle = 0, \forall \boldsymbol{x} \in S_1 \cap S_2 \tag{4}$$

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

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