

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:

$$f(\alpha \mathbf{x}) = \alpha x_n - \alpha x_1 = \alpha f(\mathbf{x})$$

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

$$f(\mathbf{x}) = x_n - x_1$$

$$f(\mathbf{y}) = y_n - y_1$$

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

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.

Non-Invertibility: For the least squares solution $\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ to be well-defined,

$\mathbf{X}^T \mathbf{X}$ must be invertible. However, with linearly dependent columns, $\mathbf{X}^T \mathbf{X}$ is singular.

Infinite Solutions: The least squares approach cannot uniquely identify β vector since there exists infinitely many solutions.

This will make class of affine functions is too large to search f.

(b) We know it should satisfy:

$$\begin{aligned}\nabla_{\beta}(\|\mathbf{X}\beta - \mathbf{y}\|_2^2 + \lambda\|\beta\|_2^2) &= 2\mathbf{X}^T(\mathbf{X}\beta - \mathbf{y}) + 2\lambda\beta = 0 \\ (\mathbf{X}^T \mathbf{X} + \lambda\mathbf{I})\beta &= \mathbf{X}^T \mathbf{y}\end{aligned}$$

Provided $\lambda > 0$, $\mathbf{X}^T \mathbf{X} + \lambda\mathbf{I}$ is invertible,

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$$

Given $\mathbf{x}^{(1)} = 2\mathbf{x}^{(2)}$, the matrix $\mathbf{X}^T \mathbf{X}$ has the form, where $a = \|\mathbf{x}^{(2)}\|_2^2$:

$$\begin{aligned}\mathbf{X}^T \mathbf{X} &= \begin{bmatrix} 4a & 2a \\ 2a & a \end{bmatrix} \\ \mathbf{X}^T \mathbf{X} + \lambda\mathbf{I} &= \begin{bmatrix} 4a + \lambda & 2a \\ 2a & a + \lambda \end{bmatrix}\end{aligned}$$

We know (give α as a constant):

$$(\mathbf{X}^T \mathbf{X} + \lambda\mathbf{I})^{-1} = \alpha \begin{bmatrix} a + \lambda & -2a \\ -2a & 4a + \lambda \end{bmatrix}$$

$$\text{Assume } (\mathbf{X}^T \mathbf{X} + \lambda\mathbf{I})^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \mathbf{X}^T = \begin{bmatrix} 2x_1 & 2x_2 \\ x_1 & x_2 \end{bmatrix}$$

$$(\mathbf{X}^T \mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}^T = \begin{bmatrix} 2ax_1 + bx_1 & 2ax_2 + bx_2 \\ 2cx_1 + dx_1 & 2cx_2 + dx_2 \end{bmatrix}$$

$$(\mathbf{X}^T \mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} = \begin{bmatrix} 2ax_1 + bx_1 & 2ax_2 + bx_2 \\ 2cx_1 + dx_1 & 2cx_2 + dx_2 \end{bmatrix} \mathbf{y} = \begin{bmatrix} (2ax_1 + bx_1)y_1 + (2ax_2 + bx_2)y_2 \\ (2cx_1 + dx_1)y_1 + (2cx_2 + dx_2)y_2 \end{bmatrix}$$

we know that $\frac{2a+b}{2c+d} = \frac{2a+2\lambda-2a}{-4a+4a+\lambda} = \frac{2\lambda}{\lambda} = 2$

It is obvious that $\hat{\beta}_1/\hat{\beta}_2 = 2$.

(c) As λ increasing, the matrix $\mathbf{X}^T \mathbf{X} + \lambda\mathbf{I}$ becomes increasingly well-conditioned and easier to invert. By adding $\lambda\mathbf{I}$, ridge regression reduces the influence of the linear dependency in \mathbf{X} . The regularization term $\lambda\|\beta\|_2^2$ effectively penalizes large values in β , which helps in controlling variance and provides a unique solution despite \mathbf{X} being rank-deficient. As $\lambda \rightarrow 0$, the ridge solution approaches the least squares solution, potentially reintroducing instability due to multicollinearity. As $\lambda \rightarrow \infty$, the solution $\hat{\beta}$ shrinks toward zero, prioritizing stability but at the cost of increasing bias. Therefore, choosing an appropriate λ balances stability and accuracy.

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:

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

Answer

Since $\forall \mathbf{a} \in H$, we can decompose it as:

$$\mathbf{a} = \mathbf{a}_s + \sum_{i=1}^N c_i \mathbf{x}_i, \text{ where } c = \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} \in \mathbb{R}^N$$

and \mathbf{a}_s satisfies $\langle \mathbf{a}_s, \mathbf{x}_i \rangle = 0, i = 1, \dots, N$

Then the objective function in (K-SVM) is:

$$\begin{aligned} & \sum_{i=1}^N h(y_i(\langle \mathbf{a}, \mathbf{x}_i \rangle + b) - 1) + \lambda \|\mathbf{a}\|_2^2 \\ &= \sum_{i=1}^N h(y_i(\langle \mathbf{a}_s + \sum_{j=1}^N c_j \mathbf{x}_j, \mathbf{x}_i \rangle + b) - 1) + \lambda \|\mathbf{a}\|_2^2 \\ &= \sum_{i=1}^N h(y_i(\sum_{j=1}^N c_j \langle \mathbf{x}_j, \mathbf{x}_i \rangle + b) - 1) + \lambda \|\mathbf{a}\|_2^2 \end{aligned}$$

Define $\xi_i = 1 - y_i(\sum_{j=1}^N c_j \langle \mathbf{x}_j, \mathbf{x}_i \rangle + b) = \max\{0, 1 - y_i(\langle \mathbf{a}, \mathbf{x}_i \rangle + b)\}$ s.t. we have to ensure:

$$\begin{aligned} \xi_i &\geq 1 - y_i(\langle \mathbf{a}, \mathbf{x}_i \rangle + b) \\ \xi_i &\geq 0 \end{aligned}$$

We know the problem becomes:

$$\min_{\mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}, \boldsymbol{\xi} \in \mathbb{R}^N} \sum_{i=1}^N \xi_i + \lambda \|\mathbf{a}\|_2^2$$

So we can conclude that solving the above soft-SVM is equivalent to solving the following problem:

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

Question 5:

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

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

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.

Answer

(a)

$$\begin{aligned} & \langle \mathbf{a}, (1+t)\mathbf{z} - t\mathbf{x} \rangle \\ &= (1+t)\langle \mathbf{a}, \mathbf{z} \rangle - t\langle \mathbf{a}, \mathbf{x} \rangle \\ &= (1+t)b - tb \\ &= b \end{aligned}$$

There \mathbf{a} can be $\mathbf{a}_1, \mathbf{a}_2$ (correspondingly b should be b_1, b_2).

We can conclude that $(1+t)\mathbf{z} - t\mathbf{x} \in S_1 \cap S_2$ for any $t \in \mathbb{R}$

(b) From (a) we know:

$$\mathbf{z} + t(\mathbf{x} - \mathbf{z}) \in S_1 \cap S_2$$

since \mathbf{z} is the $\min_{\mathbf{x} \in S_1 \cap S_2} \|\mathbf{x} - \mathbf{y}\|$

$$\begin{aligned} \|\mathbf{z} - \mathbf{y}\|^2 &\leq \|\mathbf{z} + t(\mathbf{x} - \mathbf{z}) - \mathbf{y}\|^2 \\ &= \|(\mathbf{z} - \mathbf{y}) + t(\mathbf{x} - \mathbf{z})\|^2 \\ &= \|\mathbf{z} - \mathbf{y}\|^2 + t^2\|\mathbf{x} - \mathbf{z}\|^2 + 2t\langle \mathbf{z} - \mathbf{y}, \mathbf{x} - \mathbf{z} \rangle \end{aligned}$$

Proof: If \mathbf{z} is a solution. then: Obviously, $\mathbf{z} \in S_1 \cap S_2$,

$$\begin{aligned} 2t\langle \mathbf{z} - \mathbf{y}, \mathbf{x} - \mathbf{z} \rangle &\geq -t^2\|\mathbf{x} - \mathbf{z}\|^2 \\ \langle \mathbf{z} - \mathbf{y}, \mathbf{x} - \mathbf{z} \rangle &\geq -\frac{t}{2}\|\mathbf{x} - \mathbf{z}\|^2, \text{ if } t > 0 \\ \langle \mathbf{z} - \mathbf{y}, \mathbf{x} - \mathbf{z} \rangle &\geq 0, \text{ let } t \rightarrow 0_+ \\ \langle \mathbf{z} - \mathbf{y}, \mathbf{x} - \mathbf{z} \rangle &\leq -\frac{t}{2}\|\mathbf{x} - \mathbf{z}\|^2, \text{ if } t < 0 \\ \langle \mathbf{z} - \mathbf{y}, \mathbf{x} - \mathbf{z} \rangle &\leq 0, \text{ let } t \rightarrow 0_- \\ \langle \mathbf{z} - \mathbf{y}, \mathbf{x} - \mathbf{z} \rangle &= 0 \text{ together} \end{aligned}$$

Proof: If $z \in S_1 \cap S_2$, and $\langle z - \mathbf{y}, \mathbf{x} - \mathbf{z} \rangle = 0, \forall \mathbf{x} \in S_1 \cap S_2$

$$\begin{aligned}
\|\mathbf{x} - \mathbf{y}\|^2 &= \|(\mathbf{x} - \mathbf{z}) + (\mathbf{z} - \mathbf{y})\|^2 \\
&= \|\mathbf{x} - \mathbf{z}\|^2 + \|\mathbf{z} - \mathbf{y}\|^2 + 2\langle \mathbf{x} - \mathbf{z}, \mathbf{z} - \mathbf{y} \rangle \\
&= \|\mathbf{x} - \mathbf{z}\|^2 + \|\mathbf{z} - \mathbf{y}\|^2 \\
&\geq \|\mathbf{z} - \mathbf{y}\|^2 \\
\text{so, } \mathbf{z} &= \operatorname{argmin}_{\mathbf{x} \in S_1 \cap S_2} \|\mathbf{x} - \mathbf{y}\|
\end{aligned}$$

(c) Denote $\mathbf{w}_1 = \frac{1}{\|\mathbf{a}_1\|} \mathbf{a}_1$, and $\mathbf{w}_2 = \frac{1}{\|\mathbf{w}'_2\|} \mathbf{w}'_2$, where $\mathbf{w}'_2 = \mathbf{a}_2 - \frac{\langle \mathbf{a}_2, \mathbf{a}_1 \rangle}{\langle \mathbf{a}_1, \mathbf{a}_1 \rangle} \mathbf{a}_1$

$$\mathbf{z} = \mathbf{y} - \sum_{i=1}^2 \langle \mathbf{y}, \mathbf{w}_i \rangle \mathbf{w}_i$$

$$\begin{aligned}
\langle \mathbf{a}_1, \mathbf{w}_1 \rangle &= \frac{\langle \mathbf{a}_1, \mathbf{a}_1 \rangle}{\|\mathbf{a}_1\|} \\
\langle \mathbf{a}_1, \mathbf{w}_2 \rangle &= 0 \\
\langle \mathbf{a}_1, \mathbf{z} \rangle &= 0 \text{ we can prove according to above equation}
\end{aligned}$$

Similarly, we can prove this is true, about $\langle \mathbf{a}_2, \mathbf{z} \rangle = 0$. And, we can verify that.

$$\langle \mathbf{z} - \mathbf{y}, \mathbf{x} - \mathbf{z} \rangle = \sum_{i=1}^2 \langle \mathbf{y}, \mathbf{w}_i \rangle \langle \mathbf{w}_i, \mathbf{x} - \mathbf{z} \rangle = 0$$

In conclusion, $\mathbf{w}_1 = \frac{1}{\|\mathbf{a}_1\|} \mathbf{a}_1$, and $\mathbf{w}_2 = \frac{1}{\|\mathbf{w}'_2\|} \mathbf{w}'_2$, where $\mathbf{w}'_2 = \mathbf{a}_2 - \frac{\langle \mathbf{a}_2, \mathbf{a}_1 \rangle}{\langle \mathbf{a}_1, \mathbf{a}_1 \rangle} \mathbf{a}_1$

$$\mathbf{z} = \mathbf{y} - \sum_{i=1}^2 \langle \mathbf{y}, \mathbf{w}_i \rangle \mathbf{w}_i$$

is a solution of (3)

(d) We know that: $\langle \mathbf{z} - \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle = 0, \forall \mathbf{x} \in S_1 \cap S_2$

Proof: Suppose we have 2 solutions $\mathbf{z}_1, \mathbf{z}_2$, and $\mathbf{z}_1, \mathbf{z}_2 \in S_1 \cap S_2$

$$\begin{aligned}
\langle \mathbf{z}_1 - \mathbf{y}, \mathbf{z}_2 - \mathbf{z}_1 \rangle &= 0 \\
\langle \mathbf{z}_2 - \mathbf{y}, \mathbf{z}_1 - \mathbf{z}_2 \rangle &= 0 \\
\langle \mathbf{z}_1 - \mathbf{y}, \mathbf{z}_2 - \mathbf{z}_1 \rangle + \langle \mathbf{z}_2 - \mathbf{y}, \mathbf{z}_1 - \mathbf{z}_2 \rangle &= 0 \\
\langle \mathbf{z}_1 - \mathbf{y} + \mathbf{y} - \mathbf{z}_2, \mathbf{z}_2 - \mathbf{z}_1 \rangle &= 0 \\
\langle \mathbf{z}_1 - \mathbf{z}_2, \mathbf{z}_2 - \mathbf{z}_1 \rangle &= 0 \\
-\|\mathbf{z}_1 - \mathbf{z}_2\|^2 &= 0 \implies \mathbf{z}_1 = \mathbf{z}_2
\end{aligned}$$