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:

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

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

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

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

✓

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 $\sum_{i=1,i\neq3}^{8} a_ix_i > 0$ and $b > 0$. Thus, we can not ensure $y = \sum_{i=1,i\neq3}^{8} a_ix_i + b + a_3b_3 > 0$.

(b) True.

From the condition, we can deduce that $y = \sum_{i=1,i\neq2}^{8} a_ix_i + b$, 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 $\boldsymbol{X} \in \mathbb{R}^{2\times2}$ 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:

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

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:

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

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

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

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

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

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

Infinite Solutions: The least squares approach cannot uniquely identify $\boldsymbol{\beta}$ 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:

$$\nabla_{\boldsymbol{\beta}}(\|\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y}\|_2^2+\lambda\|\boldsymbol{\beta}\|_2^2) = 2\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{y})+2\lambda\boldsymbol{\beta}=0$$
$$(\boldsymbol{X}^T\boldsymbol{X}+\lambda\boldsymbol{I})\boldsymbol{\beta}=\boldsymbol{X}^T\boldsymbol{y}$$

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

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

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

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

We know (give α as a constant):

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

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

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

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

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:

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

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:

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

Answer

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

$$\boldsymbol{a} = \boldsymbol{a}_s + \sum_{i=1}^{N} c_i \boldsymbol{x}_i, \text{ where } c = \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix} \in \mathbb{R}^N$$
$$\text{and } \boldsymbol{a}_s \text{ satisfies } \langle \boldsymbol{a}_s, \boldsymbol{x}_i\rangle = 0, i = 1, \cdots, N$$

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

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

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

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

We know the problem becomes:

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

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

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

Question 5:

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

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

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.

Answer

(a)

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

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

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

(b) From(a) we know:

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

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

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

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

$$\begin{aligned}
2t\langle \boldsymbol{z} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{z} \rangle &\ge -t^2\|\boldsymbol{x} - \boldsymbol{z}\|^2 \\
\langle \boldsymbol{z} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{z} \rangle &\ge -\frac{t}{2}\|\boldsymbol{x} - \boldsymbol{z}\|^2, \text{ if } t > 0 \\
\langle \boldsymbol{z} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{z} \rangle &\ge 0, \text{ let } t \to 0_+ \\
\langle \boldsymbol{z} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{z} \rangle &\le -\frac{t}{2}\|\boldsymbol{x} - \boldsymbol{z}\|^2, \text{ if } t < 0 \\
\langle \boldsymbol{z} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{z} \rangle &\le 0, \text{ let } t \to 0_- \\
\langle \boldsymbol{z} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{z} \rangle &= 0 \text{ together}
\end{aligned}$$

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

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

✓

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

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

$$\begin{aligned}
\langle \boldsymbol{a}_1, \boldsymbol{w}_1 \rangle &= \frac{\langle \boldsymbol{a}_1, \boldsymbol{a}_1 \rangle}{\|\boldsymbol{a_1}\|} \\
\langle \boldsymbol{a}_1, \boldsymbol{w}_2 \rangle &= 0 \\
\langle \boldsymbol{a}_1, \boldsymbol{z} \rangle &= 0 \text{ we can prove according to above equation}
\end{aligned}$$

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

-6

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

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

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

✗

is a solution of (3)

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

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

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

✓