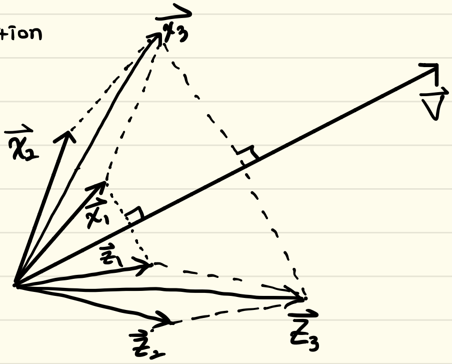
기계 인공지능 Mid sol

Problem 1

(a) Yes. Let f be the reflection transformation.

To prove this transformation is linear, we need to show
$$\begin{cases} \text{① superposition} & f(x_1+x_2)=f(x_1)+f(x_2) \\ \text{② homogeneity} & f(ax_1)=af(x_1) \end{cases}$$

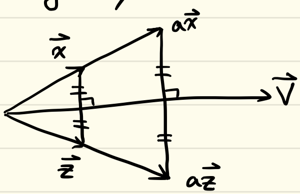
① superposition



Suppose
$$\left(\begin{array}{ll} f(\vec{x}_1)=\vec{z}_1, & f(\vec{x}_2)=\vec{z}_2 \\ f(\vec{x}_3)=\vec{z}_3, & \vec{x}_3=\vec{x}_1+\vec{x}_2 \\ \vec{z}_1+\vec{z}_2=\vec{z}_3 & \end{array} \right)$$

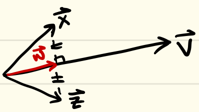
So, $f(\vec{x}_1+\vec{x}_2)=f(\vec{x}_3)=\vec{z}_3=\vec{z}_1+\vec{z}_2=f(\vec{x}_1)+f(\vec{x}_2)$

② Homogeneity



$f(a\vec{x})=a\vec{z}=af(\vec{x})$

(b)



Suppose that $\vec{\omega}$ is a projection vector of $\vec{x}$ onto $\vec{V}$

Then, $\vec{\omega}=\dfrac{\vec{V}^T\vec{x}}{\vec{V}^T\vec{V}}\vec{V}=\dfrac{\vec{V}\vec{V}^T}{\vec{V}^T\vec{V}}\vec{x}$

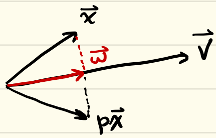
$\vec{z}=\vec{x}-2(\vec{\omega}-\vec{x})=2\vec{\omega}-\vec{x}$

$=\left(2\dfrac{\vec{V}\vec{V}^T}{\vec{V}^T\vec{V}}-1\right)\vec{x}$

so, $M=2\dfrac{\vec{V}\vec{V}^T}{\vec{V}^T\vec{V}}-1$

(c) From geometric interpretation, we know that $\vec{w}$ and $\vec{x}-\vec{w}$ are eigenvectors of the reflection transformation, and corresponding eigenvalues are 1 and -1, respectively.

Let P be a reflection matrix. Then,



$$P\vec{x} = \vec{w} + (-1)(\vec{x}-\vec{w}) = 2\vec{w} - \vec{x}$$
$$= \left(2\frac{vv^T}{v^Tv} - I\right)\vec{x}$$

So, $M = 2\dfrac{vv^T}{v^Tv} - I$

(d) Let $v = [1 \quad 1]^T$

Then from the equation for M,

$$M = 2\cdot\frac{\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix}}{\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}} - I = \frac{2\cdot\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}}{2} - \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$$

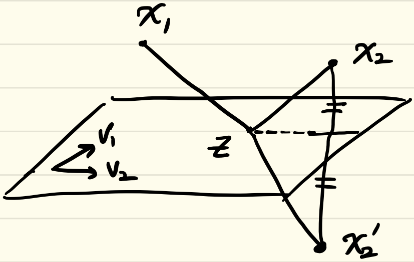
$$Mx = \lambda x \Rightarrow (M-\lambda I)x = 0$$

$$\det\begin{bmatrix}-\lambda & 1\\1 & -\lambda\end{bmatrix} = \lambda^2 - 1 = 0 \quad \therefore \lambda = \pm 1$$

$$\lambda = +1 \Rightarrow \begin{bmatrix}-1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = 0 \rightarrow u_1 = \begin{bmatrix}1\\1\end{bmatrix}$$

$$\lambda = -1 \Rightarrow \begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = 0 \rightarrow u_2 = \begin{bmatrix}1\\-1\end{bmatrix}$$

Problem 2



$x_2' = f(x_2)$
where f is a reflection transformation w.r.t $V = [V_1 \ V_2]$

Since the x_2' is a result of reflection transformation of x_2, the length $\overline{x_2 z} = \overline{x_2' z}$.
So, this problem can be rewrite as

$$d = \min_z \left(\|x_1 - z\| + \|x_2' - z\| \right)$$

the length d is minimized when x_1, x_2', z are on the same line.

So,

$$d = \min \left[\|x_1 - z\|_2 + \|x_2 - z\|_2 \right]$$
$$= \|x_1 - x_2'\|_2$$

$$x_2' = M x_2 = \left(2\frac{VV^T}{V^TV} - I\right) x_2$$ from problem 1.

Then, $d = \left\| x_1 - \left(2\frac{VV^T}{V^TV} - I\right) x_2 \right\|$

Problem 3

(1) $x^T y = \begin{bmatrix} x[1] & x[2] & \cdots & x[m] \end{bmatrix} \begin{bmatrix} y[1] \\ \vdots \\ y[m] \end{bmatrix}$

$= x[1]y[1] \cdots\cdots x[m]y[m]$

by using an inner product, we can determine similarity of two signals.

As increasing similarity of two signals, the value of an inner product is increasing.

problem 4

(a) Sum of 2-norms of multiple vectors can be expressed as a 2-norm of a single vector.

$$\|A\theta - y\|_2^2 + \|\sqrt{\lambda} I\theta - 0\|_2^2 = \left\| \begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix} \theta - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|_2^2$$

$$\hat{\theta} = \left(\begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix}^T \begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix} \right)^{-1} \begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix}^T \begin{bmatrix} y \\ 0 \end{bmatrix}$$

$$= (A^T A + \lambda I_n)^{-1} A^T y$$

// done

(b) Let $J(\theta) = \|A\theta - y\|_2^2 + \lambda \|\theta\|_2^2$

A gradient descent algorithm is formulated as the following.

$$\theta \leftarrow \theta - \eta \frac{\partial J}{\partial \theta}$$

Need to compute $\frac{\partial J}{\partial \theta}$

$$g_{proj} = \frac{\partial}{\partial \theta} \|A\theta - y\|_2^2$$
$$= \frac{\partial}{\partial \theta} (A\theta - y)^T (A\theta - y)$$
$$= \frac{\partial}{\partial \theta} (\theta^T A^T A \theta - 2\theta^T A^T y - y^T y)$$
$$= 2A^T A \theta - 2A^T y$$

$$g_{reg} = \frac{\partial}{\partial \theta} \|\theta\|_2^2$$
$$= \frac{\partial}{\partial \theta} \theta^T \theta$$
$$= 2\theta$$

Thus, $\frac{\partial J}{\partial \theta} = g_{proj} + \lambda g_{reg}$

$$= 2A^T A\theta - 2A^T y + 2\theta\lambda$$

Then, the gradient descent algorithm is formulated as follows

$$\theta \leftarrow \theta - \eta (2A^T A \theta - 2A^T y + 2\theta\lambda)$$

(C)

loss function of this problem is $\|A\theta - y\|_2^2 + \lambda\|\theta\|_2^2$
and we want to minimize it,
to minimize first term $\|A\theta - y\|_2^2$, fit all the possible data points.
It causes overfitting problem.
So, we need regularization term. $\lambda\|\theta\|_2^2$.
Let see $g = g_{proj} + \lambda g_{reg}$, g should be zero to minimize loss function
$g_{reg} = 2\theta$ and if we reduce g_{reg}, (dominantly minimize the weight θ)
Then, $\|\theta\|_2 = \sqrt{\theta_1^2 + \cdots \theta_m^2}$ converge to zero when each θ_i are close to 0.

And so it has the effect of shrinking the estimates of θ_i
It has the same effect as dimension reduction,
Thus, overfitting problem can be solved.

(d) In the problem 4-(a)

$$\underbrace{(A^TA + \lambda I_n)^{-1}}_{\text{regularizer}}\underbrace{A^Ty}_{\text{projection}} \leftarrow (A^TA)^{-1}\underbrace{A^Ty}_{\text{projection}}$$

regularize term is in the denominator so it make θ converge to zero, (adjust the projection term)
From the (c) and above sentence, we can see (a) and (b) has same meaning.

Problem 5

(1) $f(x_1, x_2) = 2x_1 + 3x_2 + 1 = 0$

$\vec{w} = (2, 3)$

$x_a = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \Rightarrow d_a = \frac{g(x_a)}{\|w\|} = \frac{4+3+1}{\sqrt{4+9}} = \frac{8}{\sqrt{13}}$

$x_b = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow d_b = \frac{g(x_b)}{\|w\|} = \frac{2-3+1}{\sqrt{4+9}} = 0$

(2) $w^T x + w_0 = 0$

$\vec{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$ then, $w_1 x_1 + \cdots w_n x_n + w_0 = 0$

And the minimum γ is distance between the hyperplane and the origin

$$r = \frac{\|w_0\|}{\sqrt{w_1^2 + \cdots w_n^2}}$$

Problem 6

(1)

Only w_2 is penelized $\rightarrow$ w_1 will be dominant.
This has the same effect as maximizing the absolute value of the slope $|\frac{w_1}{w_2}|$ which implies the decision boundry
$\Rightarrow$ slope should be bigger than L_1 slope.

Only L_3 has a bigger slope with its sign unchanged.

(2) (a)

As we increase the regularization prameter λ, we need to reduce $|w_1|$ and $|w_2|$

more dominant one should be minimized prior to the other.
From figure 5, we can see that
x_1 axis is more critical in classification ($x_2 \cong 0$)
Since x_2 axis has little effect, we send w_1 to zero before w_2.
than, w_2 also goes to zero.