1

(a)

For $\vec{x_1}$ and $\vec{x_2}$ on the hyperplane, we have the equation as follows:

$$g(\vec{x_1}) = g(\vec{x_2}) \iff w^T \cdot \vec{x_1} + w_0 = w^T \vec{x_2} + w_0$$

$$\iff w^T(\vec{x_1} - \vec{x_2}) = 0$$

Since the weight vector is perpendicular to any vector on the hyperplane, the weight vector is normal to the hyperplane.

(b) We assume $\vec{x} = \vec{x_1} + a\vec{w}$

The distance d from H to $\vec{x}$ is:

$$d = \frac{g(\vec{x})}{\|\vec{w}\|} = \frac{g(\vec{x_1} + a\vec{w})}{\|\vec{w}\|} = \frac{\vec{w}^T(\vec{x_1} + a\vec{w}) + \vec{w_0}}{\|\vec{w}\|} = \frac{\vec{w}^T\vec{x_1} + a\vec{w}^T\vec{w} + \vec{w_0}}{\|\vec{w}\|}$$

Since $\vec{x_1}$ is on H, we have $g(\vec{x_1}) = \vec{w}^T\vec{x_1} + \vec{w_0} = 0$

So $d = \frac{a\|\vec{w}\|^2}{\|\vec{w}\|} = a\|\vec{w}\|$

So when $a > 0$, $a\vec{w}$ points to the positive side of H and we have $d = a\|\vec{w}\| > 0$ as $\|\vec{w}\|$ is positive.

Q.E.D

(c) Similar with the proof above, we assume a point $\vec{x_p}$ on plane H and a point $\vec{x}$ on the positive side of H.

So we have $\vec{x} = \vec{x_p} + r\frac{\vec{w^+}}{\|\vec{w^+}\|}$

Since it is augmented space and for $\vec{x_p}$,

$$g(\vec{x_p}) = w^{+} \cdot x_p^+ = 0$$

So $g(\vec{x^+}) = g(\vec{x_p^+} + r\frac{\vec{w^+}}{\|\vec{w^+}\|}) = \vec{w^+} \cdot \vec{x_p^+} + r \cdot \frac{\vec{w^+} \cdot \vec{w^+}}{\|\vec{w^+}\|} = r \cdot \|w^+\|$

$$r = \frac{g(\vec{x^+})}{\|w^+\|}$$

Q.E.D

(d) Similar with the proof in feature space:

Let $\vec{w_p}$ is a point on H thus we can express the $\vec{w}$ on the positive side as: $\vec{w}^+ = \vec{w_p}^+ + r\frac{\vec{x}^+}{\|\vec{x}\|}$

So $g_{\vec{x}^+}(\vec{w}^+) = g(\vec{w_p}^+ + r\frac{\vec{x}^+}{\|\vec{x}^+\|}) = \vec{w_p}^+ \cdot \vec{x}^+ + r\frac{\vec{x}^+ \cdot \vec{x}^+}{\|\vec{x}^+\|}$

Since for points on plane H, we have $g_{\vec{x}^+}(\vec{w}^+) = \vec{w_p}^+ \cdot \vec{x}^+ = 0$

$\therefore r = \frac{g_{\vec{x}^+}(\vec{w}^+)}{\|\vec{x}^+\|}$

Q.E.D

EE559 Homework 3 (week 4)

Jingquan Yan

USC ID: 1071912676

Email: jingquan@usc.edu

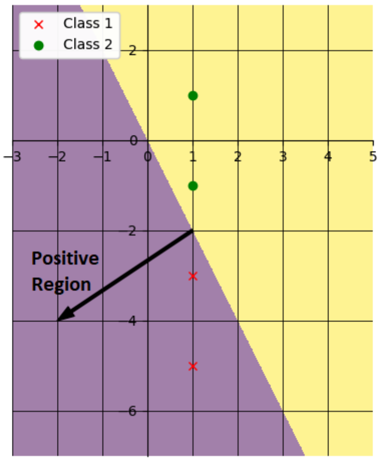
EE559 repository: [Github]()

---

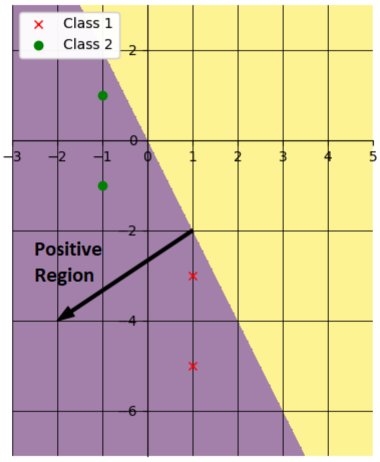
2 :

Solution:

1. Data set and linear decision boundary is as follows:

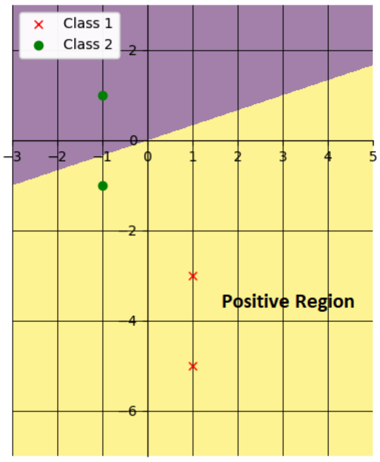


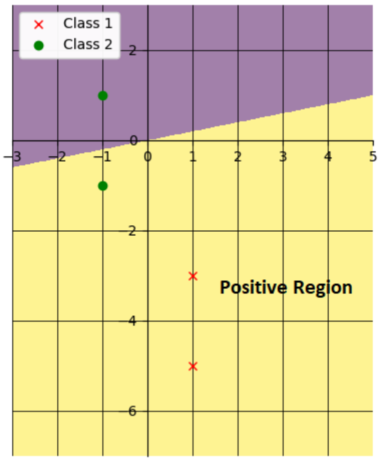
2. Reflected data points and the same decision boundary are as follow:

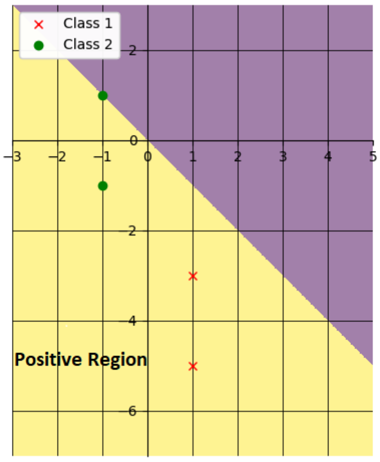


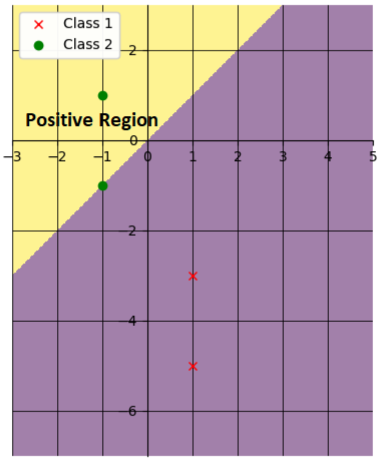
   The same decision boundary can't not classify the reflected points correctly. However, since the reflected data points are at the same side of the boundary, this fact proves the linear separability of the 2-class points.

3. The respective positive regions of each reflected data points are as follow:

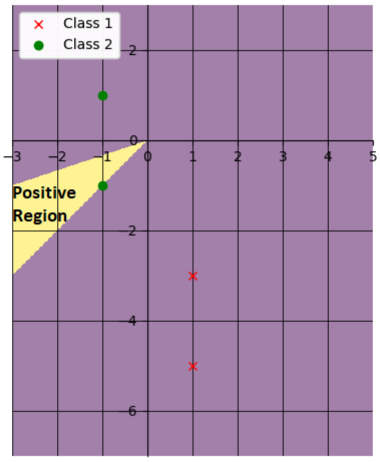




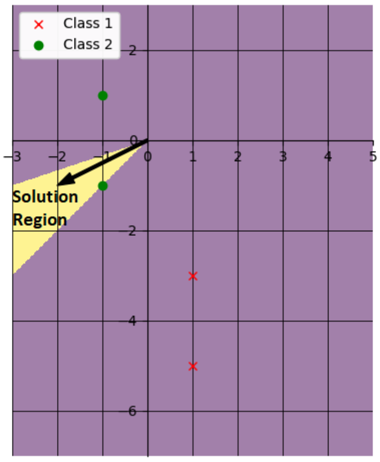




The final **solution region** is:



4. The weight vector of part (a) is:



It is obvious that the previous weight vector is in the solution area thus it can successfully separate the 2-class points.

3. (a) From definition we have: $\nabla_{\underline{x}} f(\underline{x}) = \frac{\partial f(\underline{x})}{\partial \underline{x}} = \left[\frac{\partial f(\underline{x})}{\partial x_1}, \dots, \frac{\partial f(\underline{x})}{\partial x_n}\right]^T$

So for $\nabla_{\underline{x}} f[p(\underline{x})] = \left[\frac{\partial f(p(\underline{x}))}{\partial x_1}, \dots, \frac{\partial f(p(\underline{x}))}{\partial x_n}\right]^T$

Since $\frac{\partial f(p(\underline{x}))}{\partial x_i} = \frac{d(f(p))}{dp} \cdot \frac{\partial p(\underline{x})}{\partial x_i}$

So $\nabla_{\underline{x}} f[p(\underline{x})] = \left[\frac{d(f(p))}{dp}\frac{\partial p(\underline{x})}{\partial x_1}, \dots, \frac{d(f(p))}{dp}\frac{\partial p(\underline{x})}{\partial x_n}\right] = \left[\frac{d(f(p))}{dp}\right] \nabla_{\underline{x}} p(\underline{x})$

Q.E.D

(b) For quadratic form $f(\underline{x}) = \underline{x}^T M \underline{x}$, $\frac{\partial f(\underline{x})}{\partial \underline{x}} = (M + M^T)\underline{x}$

So when $M = I$, we have: $\frac{\partial f(\underline{x})}{\partial \underline{x}} = (I + I^T)\underline{x} = 2 \cdot \underline{x}$

$\therefore \nabla_{\underline{x}}(\underline{x}^T \underline{x}) = 2\underline{x}$

(c) $\nabla_{\underline{x}}(\underline{x}^T \underline{x}) = \nabla_{\underline{x}}(\underline{x}^T I \underline{x}) = \left(\frac{\partial}{\partial x_1}\sum_{i,j=1}^{n} a_{ij} x_i x_j, \dots, \frac{\partial}{\partial x_n}\sum_{i,j=1}^{n} a_{ij} x_i x_j\right)^T$

$$= \begin{pmatrix} 2a_{11}x_1 + (a_{12}+a_{21})x_2 + \cdots + (a_{1n}+a_{n1})x_n \\ \vdots \\ (a_{1n}+a_{n1})x_1 + (a_{n2}+a_{2n})x_2 + \cdots + 2a_{nn}x_n \end{pmatrix}$$

for I matrix, $a_{ij} = \begin{cases} 0 & \text{when } i \neq j \\ 1 & \text{when } i = j \end{cases}$

So $\nabla_{\underline{x}}(\underline{x}^T \underline{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{pmatrix} = 2\underline{x}$

(d) $\nabla_{\underline{x}}\left[(\underline{x}^T \underline{x})^3\right] = \frac{d(\underline{x}^T\underline{x})^3}{d(\underline{x}^T\underline{x})} \cdot \nabla_{\underline{x}}(\underline{x}^T\underline{x}) = 3(\underline{x}^T\underline{x})^2 \cdot 2\underline{x} = 6(\underline{x}^T\underline{x})^2\underline{x}$

4

(a) $p = \underline{w}^T \cdot \underline{w}$ and let $f = \sqrt{p}$

So $\nabla_{\underline{w}} \|\underline{w}\|_2 = \frac{d\sqrt{p}}{dp} \cdot \nabla_{\underline{w}}\, \underline{w}^T \underline{w}$

$$= \frac{1}{2\sqrt{\underline{w}^T\underline{w}}} \cdot 2\underline{w}$$

$$= \frac{1}{\sqrt{\underline{w}^T\underline{w}}} \cdot \underline{w}$$

(b) Let $p = (M\underline{w} - \underline{b})^T (M\underline{w} - \underline{b})$ and denote $M\underline{w} - \underline{b} = \underline{z}$

So $p = z^T z$

$$\nabla_{\underline{w}} \|M\underline{w} - \underline{b}\|_2 = \frac{d\sqrt{p}}{dp} \cdot \nabla_{\underline{w}}\, z^T \cdot z = \frac{1}{2\sqrt{z^T z}} \cdot \nabla_{\underline{w}} \cdot (M\underline{w} - \underline{b})^T \cdot (M\underline{w} - \underline{b})$$

$$= \frac{1}{2\sqrt{z^T z}} \cdot 2z \cdot \frac{\partial (M\underline{w} - \underline{b})}{\partial \underline{w}}$$

$$= \frac{z}{\sqrt{z^T z}} M^T$$

$$= \frac{(M\underline{w} - \underline{b}) \cdot M^T}{\sqrt{(M\underline{w} - \underline{b})^T \cdot (M\underline{w} - \underline{b})}}$$

Extra Credit

Part 1: Total linear separable $\rightarrow$ Linear separable

proof: It is equal to prove the contrapositive statement:

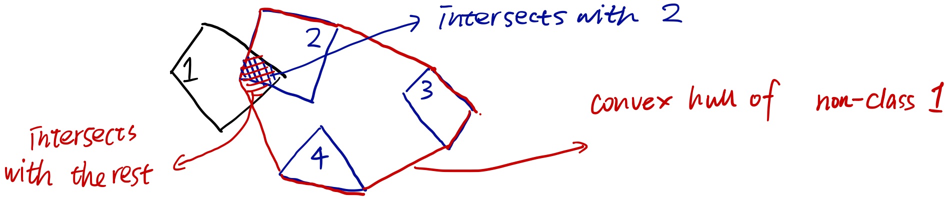
Not linear separable $\rightarrow$ Not total linear separable

From the conclusion in HW2 problem 4 we have:

Not linear separable $\Longleftrightarrow$ Convex hulls intersect

We assume convex hulls of class 1 and 2 intersect.

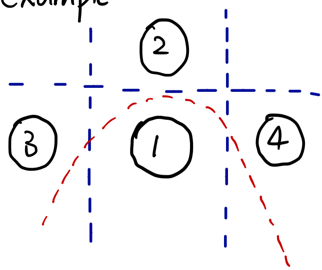
So it's obvious that the convex hull 1 intersects with the convex hull that contains the rest classes $(2, 3, \dots, n)$.



Now we can say convex hull is not linearly separable with the convex hull that contains the rest classes' elements. Thus OVR is not applicable for class 1 and it's not total linear separable. So the contrapositive statement is true and so is the original statement.

Part 2: Linear Separable $\not\rightarrow$ Total linear separable

counter example:



① is linear separable with ② ③ and ④ respectively.

But ① is not linear separable with the rest (② and ③ and ④). Thus this is not total linear separable