Problem 1

Prove Eqn (6) in Lecture 5, i.e.,

$$F1 = \frac{\text{TP}}{\text{TP} + \frac{\text{FN}+\text{FP}}{2}}$$

Precision $P = \frac{TP}{TP+FP}$ & Recall $R = \frac{TP}{TP+FN}$

$$F1 = \frac{2 \times P \times R}{P + R}$$

by substitution:

$$F1 = \frac{2 \cdot \frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}}$$

$$= \frac{\frac{2 \cdot (TP)^2}{(TP+FP)(TP+FN)}}{\frac{TP}{TP+FP} \cdot \frac{TP+FN}{TP+FN} + \frac{TP}{TP+FN} \cdot \frac{TP+FP}{TP+FP}}$$

$$= \frac{\frac{2(TP)(TP)}{(TP+FP)(TP+FN)}}{\frac{TP(TP+FN) + TP(TP+FP)}{(TP+FP)(TP+FN)}}$$

$$= \frac{2(TP)(TP)}{\cancel{(TP+FP)}\cancel{(TP+FN)}} \cdot \frac{\cancel{(TP+FP)}\cancel{(TP+FN)}}{TP((TP+FN)+(TP+FP))}$$

$$= \frac{2(\cancel{TP})(TP)}{\cancel{TP}((TP+FN)+(TP+FP))}$$

$$F1 = \frac{TP}{TP + \frac{FN+FP}{2}}$$

Problem 2

Consider the linear decision function

$$s(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$$

as defined in Lecture 6, which deals with binary classification problems.

(a) Show that $\boldsymbol{w}$ is perpendicular (orthogonal) to every vector lying within the decision boundary.

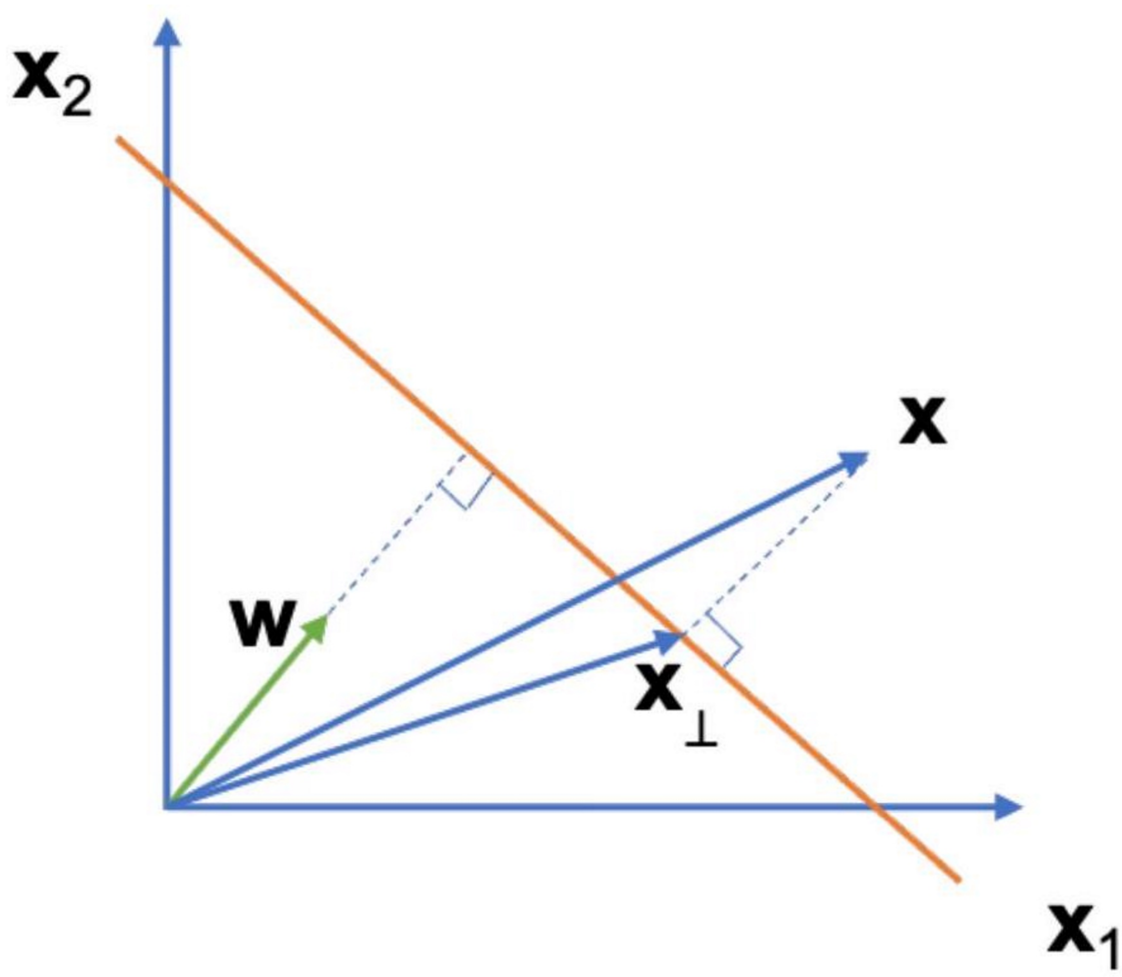
(b) Show that the distance from the origin to the decision boundary is given by

$$\frac{|b|}{||\boldsymbol{w}||_2}$$

(c) Let $\boldsymbol{x}$ be an arbitrary point. Show that the distance from $\boldsymbol{x}$ to the decision boundary is

$$\frac{|s(\boldsymbol{x})|}{||\boldsymbol{w}||_2}$$

(Hint: Use the figure below to visualize the geometry.)



(a) decision boundary:

$S(x) = W^T x + b = 0$

if X_1 & X_2 are within boundary:

$W^T X_1 + b = 0$

$W^T X_2 + b = 0$

$\therefore W^T X_1 = -b = W^T X_2$

Let $V = (X_2 - X_1)$ $\Rightarrow$ parallel to boundary

$$\therefore\ W^T V = W^T(X_2 - X_1)$$
$$= W^T X_2 - W^T X_1$$
$$= -b - (-b)$$
$$= 0$$

The dot product of orthogonal vectors is 0. Thus W is orthogonal to V, and also to every vector in the decision boundary.

(b)

The shortest path would need to satisfy:

① perpendicular to decision boundary

② parallel to W

Thus Let $V = \frac{W}{\|W\|_2}$, where V is a unit normal vector.

Let P be the intersection point between the boundary and the line from origin.

$\therefore$ length from origin to P is:

$$d = |P \cdot V|$$
$$= \frac{|W \cdot P|}{\|W\|_2}$$

$$= \frac{|W^T P|}{||W||_2}$$

because $W^T p + b = 0$

$W^T p = -b$

$$\therefore \quad d = \frac{|b|}{||W||_2}$$

(c) from part (b), we know that ① & ② must be satisfied. Then let $V = \frac{W}{||W||_2}$ & Let X_D be where the line from X intersects the boundary.

$$d = |X \cdot V|$$

$$= \frac{|W^T X|}{||W||_2}$$

we know $W^T X = S(x) - b$

$$\therefore \quad d = \frac{|S(x) - b|}{||W||_2}$$

$$d = \frac{|S(X)|}{||W||_2}$$

because b is a bias term, which doesn't depend on X

Problem 3

(a) What is t_i^2?

(b) Use Eqns. (8) and (12), and Part (a) to show Eqn. (13) in Lecture 6.

(a) t_i is simply the ith data point, with values either 1 or -1. Thus $t_i^2 = 1$.

(b) $$t_i(w^T x_i + b) = 1 \qquad (8)$$

$$w = \sum_{i=1}^{n} a_i t_i x_i \qquad (12)$$

$$b = \frac{1}{N_S} \sum_{i \in S} \left(t_i - \sum_{j=1}^{N} a_j t_j k(x_i, x_j) \right) \qquad (13)$$

$$t_i \cdot t_i(w^T x_i + b) = 1 \cdot t_i$$

$$\sum_{i \in S} t_i^2 (w^T x_i + b) = \sum_{i \in S} t_i$$

$$w^T = \sum_{i=1}^{N} a_i t_i x_i^T = \sum_{j=1}^{n} a_j t_j x_j^T$$

This is because $\sum_{i \in S}$ refers to support vectors, while w refers to all data.

$$\therefore \sum_{i \in S} \left(\left(\sum_{j=1}^{N} a_j t_j x_j^T \right) x_i + b \right) = \sum_{i \in S} t_i$$

recall $k(x_i, x_j) = x_i^T x_j$

$$\sum_{i \in S} \left(\left(\sum_{j=1}^{N} a_j t_j k(x_i, x_j) \right) + b = \sum_{i \in S} t_i$$

$$\therefore \sum_{i \in S} b + \sum_{i \in S} \left(\sum_{j=1}^{N} a_j t_j k(x_i, x_j) \right) = \sum_{i \in S} t_i$$

$$N_S b = \sum_{i \in S} t_i - \sum_{i \in S} \left(\sum_{j=1}^{N} a_j t_j k(x_i, x_j) \right)$$

$$b = \frac{1}{N_S} \sum_{i \in S} \left(t_i - \sum_{j=1}^{N} a_j t_j k(x_i, x_j) \right)$$

Problem 4 (MATH 5388 Only)

Show Eqn. (15) is the classification constraint, i.e., for any point $\boldsymbol{x}_i$, the inequality in Eqn. (15) holds.

$$t_i s(X_i) \geq 1 - \varepsilon_i, \quad 1 \leq i \leq N \qquad (15)$$

Let's look at the possible different cases: $S(X_i)$ is positive and inside boundary:

$$t_i = 1 \;\; \& \;\; S(X_i) \geq 1 \quad \leftarrow \text{misclassified}$$

$$\therefore \; \varepsilon_i = |1 - S(X_i)|$$

$$t_i s(X_i) \geq 1 - (1 - S(X_i)) = S(X_i)$$

Thus $t_i S(X_i) \geq 1 - \varepsilon, \; 1 \leq i \leq N$ held true

next case, $S(X_i)$ is negative & inside boundary:

$$t_i = -1 \;\; \& \;\; S(X_i) \leq -1$$

$$\therefore \; \varepsilon_i = |-1 - S(X_i)|$$

$$t_i S(X_i) \geq 1 - (-1 - (S(X_i)) = S(X_i) + 2$$

$$\therefore \; -S(X_i) \geq 2 + S(X_i)$$

$$0 \geq 2 + S(X_i) + S(X_i)$$

$$-1 \geq S(X_i)$$

$$\& \;\; 2 + S(X_i) \leq 1$$

$$\therefore \; t_i S(X_i) \geq 2 + S(X_i)$$

so $t_i S(X_i) \geq 1 - \varepsilon_i$ holds true again

last case, X is on margin boundary:

$$\varepsilon_i = 0$$

$$\therefore \; t_i S(X_i) \geq 1 \geq 1 - \varepsilon_i, \quad 1 \leq i \leq N$$

thus it holds here as well.

$\therefore \; t_i S(X_i) \geq 1 - \varepsilon_i, \; 1 \leq i \leq N$ holds for all possible cases of X_i.

all possible cases of X_i.