Learning & Decision Surfaces

The techniques from linear algebra developed so far allow us to precisely describe decision surfaces for binary classification problems.

Definition: ***Decision surfaces*** are (hyper)planes that separate points in a dot product space according to their classification label.

Decision Surfaces & Functions

Let us cast our classification problem into the machine learning framework:

- Let some dot product space $\mathbb{R}^n$ be our data universe with points $\overline{x} \in \mathbb{R}^n$.
- Let S be a sample set such that $S \subset \mathbb{R}^n$.
- Let $f : \mathbb{R}^n \rightarrow \{-1, +1\}$ be the target function
- Let $D = \{\langle \overline{x}, f(\overline{x})\rangle \mid \overline{x} \in S\}$ be the training set.

Compute a function $\hat{f} : \mathbb{R}^n \rightarrow \{+1, -1\}$ using D such that

$$\hat{f}(\overline{x}) \cong f(\overline{x}) \text{ for all } \overline{x} \in \mathbb{R}^n.$$

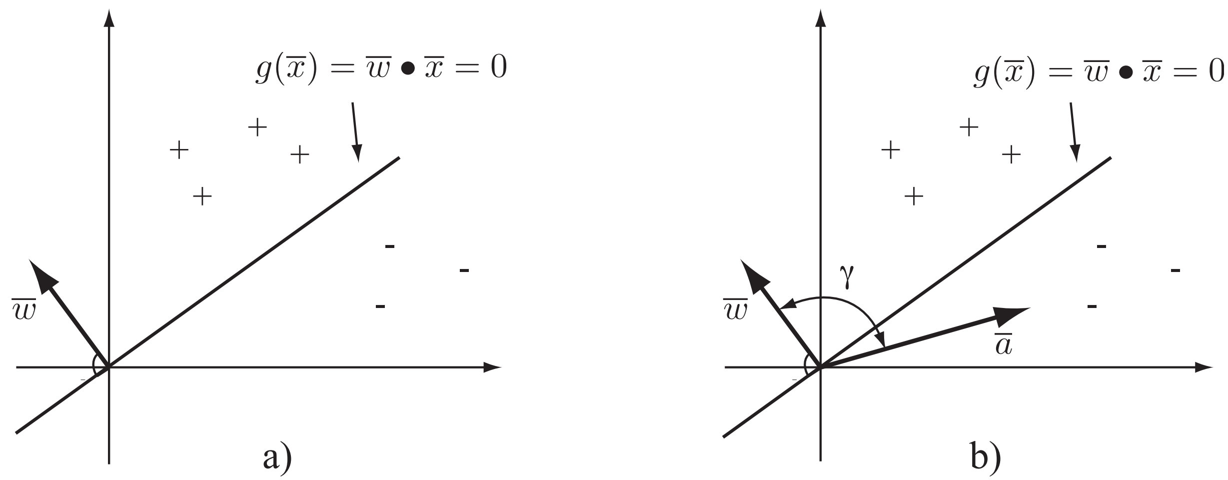
Let us assume that our training set is $D = \{\langle \overline{x}_1, y_1\rangle, \langle \overline{x}_2, y_2\rangle, \ldots, \langle \overline{x}_n, y_n\rangle\}$ with $y_i \in \{+1, -1\}$, is linearly separable. That is, we assume that we are guaranteed to find a hyperplane that separates the two classes. We also assume that we can compute a hyperplane that goes through the origin,

$$g(\overline{x}) = \overline{w} \bullet \overline{x} = 0$$

and separates the two classes. We call g a ***decision surface***.

Decision Surfaces & Functions

If we assume that our dot product space is the two dimensional real space $\mathbb{R}^2$, then we can represent a decision surface $g(\overline{x})$ as a line that goes through the origin (a):



In (b) notice that our decision surface $g(\overline{x})$ will produce positive values for points that lie above the surface ($\gamma < 90°$) and negative values for points that lie below it ($\gamma > 90°$).

Decision Surfaces & Functions

Instead of scalar positive and negative values we want our model $\hat{f}$ to compute the labels $+1$ and -1, so we use the decision surface $g(\overline{x})$ to construct our model as follows,

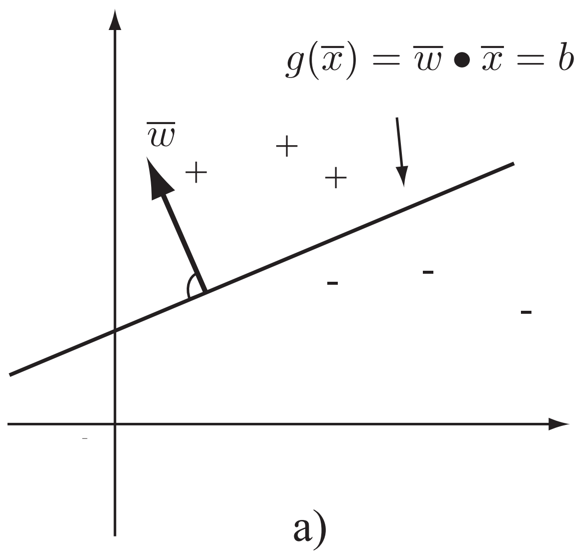
$$\hat{f}(\overline{x}) = \begin{cases} +1 & \text{if } g(\overline{x}) \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

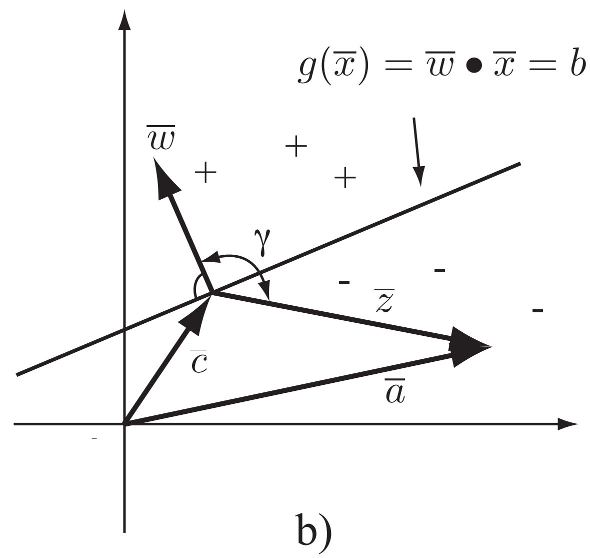
Our model $\hat{f}$ is also called a ***decision function***.

Decision Surfaces & Functions

Let us relax the requirement that our decision surface needs to go through the origin. What would our decision function look like in this case?

We can derive it as follows. Consider the following graphs.





Part (a) shows a decision function $g(\overline{x}) = \overline{w} \bullet \overline{x} = b$ with the offset term b. Part (b) shows that we can no longer simply take the angle between the position vector of the point we want to classify $\overline{a}$ and the normal vector $\overline{w}$.

Decision Surfaces & Functions

The trick here is to pick a point $\overline{c} \in g$ or

$$g(\overline{c}) = \overline{w} \bullet \overline{c} = b,$$

that can act as our origin for classification purposes. Then we can pick a $\overline{z}$ such that $\overline{a} = \overline{c} + \overline{z}$ or

$$\overline{z} = \overline{a} - \overline{c}.$$

Observe that

$$\overline{w} \bullet \overline{z} = |\overline{w}||\overline{z}| \cos \gamma$$

will give us the correct classification of point $\overline{a}$. Formally,

$$\begin{aligned} \overline{w} \bullet \overline{z} &= \overline{w} \bullet (\overline{a} - \overline{c}) \\ &= \overline{w} \bullet \overline{a} - \overline{w} \bullet \overline{c} \\ &= \overline{w} \bullet \overline{a} - b \\ &= g(\overline{a}) - b \end{aligned}$$

This means that we can compute the classification value of some point $\overline{a}$ by first applying the decision surface and then subtracting the offset term.

Decision Surfaces & Functions

We can now construct our decision function as

$$\hat{f}(\overline{x}) = \begin{cases} +1 & \text{if } g(\overline{x}) - b \geq 0, \\ -1 & \text{if } g(\overline{x}) - b < 0, \end{cases}$$

for all $\overline{x} \in \mathbb{R}^2$.

Note that decision function based on decision surfaces that run through the origin are simply a special case of the decision function above with $b = 0$.

Also note that this construction easily generalizes to arbitrarily dimension real spaces $\mathbb{R}^n$ since none of the construction depends on the dimensionality of the underlying space.