Decision Surfaces & Functions

For some decision surface $g(\overline{x}) = \overline{w} \bullet \overline{x} = b$ in an n-dimensional dot product space $\mathbb{R}^n$ we can always construct the decision function,

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

for all $\overline{x} \in \mathbb{R}^n$. Or in more compact form,

$$\boxed{\hat{f}(\overline{x}) = \operatorname{sign}(\overline{w} \bullet \overline{x} - b),}$$

with $\overline{w}, \overline{x} \in \mathbb{R}^n$, $b \in \mathbb{R}$, and

$$\operatorname{sign}(k) = \begin{cases} +1 & \text{if } k \geq 0, \\ -1 & \text{if } k < 0, \end{cases}$$

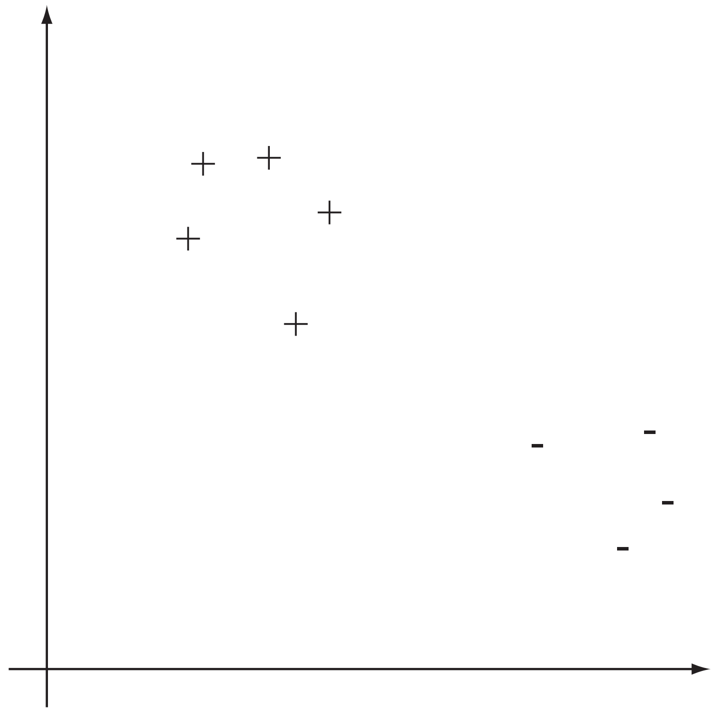
for all $k \in \mathbb{R}$.

A Simple Learning Algorithm

Let's investigate a simple algorithm that actually computes a decision surface for our training set

$$D = \{(\overline{x}_1, y_1), (\overline{x}_2, y_2), \ldots, (\overline{x}_l, y_l)\},$$

with $\overline{x}_i \in \mathbb{R}^2$ and $y_i \in \{+1, -1\}$. Here we relax the restriction that the decision surface has to go through the origin but we still assume that D is linearly separable.



A Simple Learning Algorithm

Step 1.

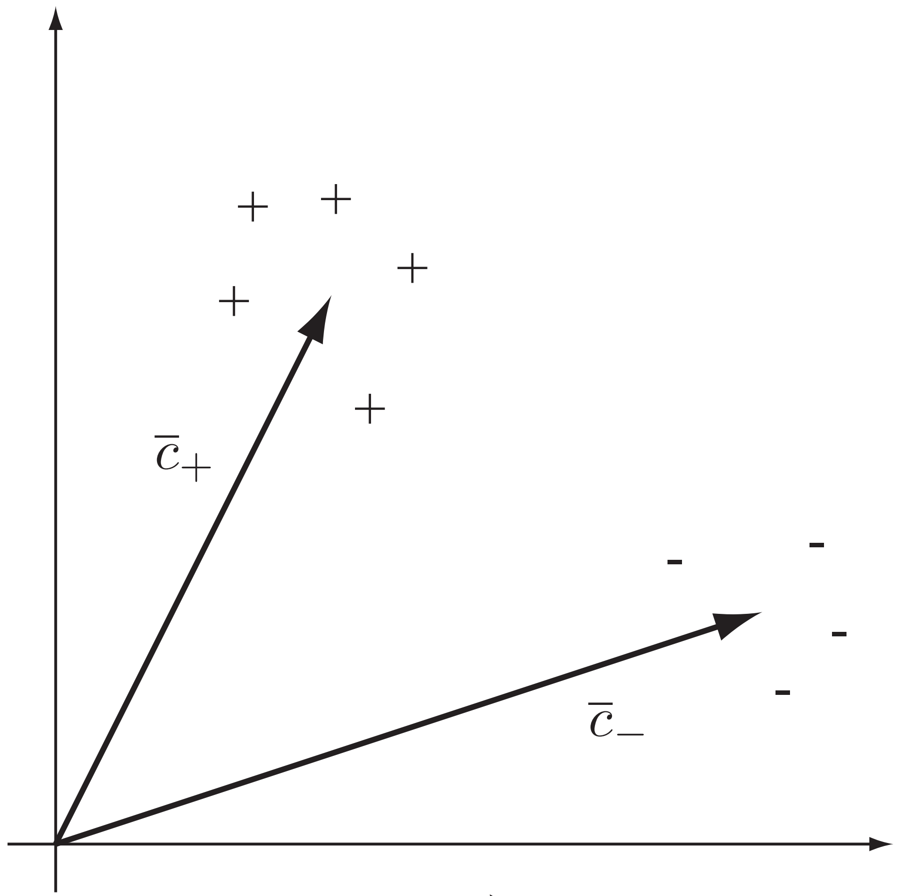
Compute the mean vectors, $\overline{c}_+$ and $\overline{c}_-$, for each class, respectively:

$$\begin{aligned} \overline{c}_+ &= \frac{1}{l_+} \sum_{(\overline{x}_i, +1) \in D} \overline{x}_i, \\ \overline{c}_- &= \frac{1}{l_-} \sum_{(\overline{x}_i, -1) \in D} \overline{x}_i, \end{aligned}$$

where

$$\begin{aligned} l_+ &= |\{(\overline{x}, y) \mid (\overline{x}, y) \in D \text{ and } y = +1\}|, \\ l_- &= |\{(\overline{x}, y) \mid (\overline{x}, y) \in D \text{ and } y = -1\}|. \end{aligned}$$

A Simple Learning Algorithm



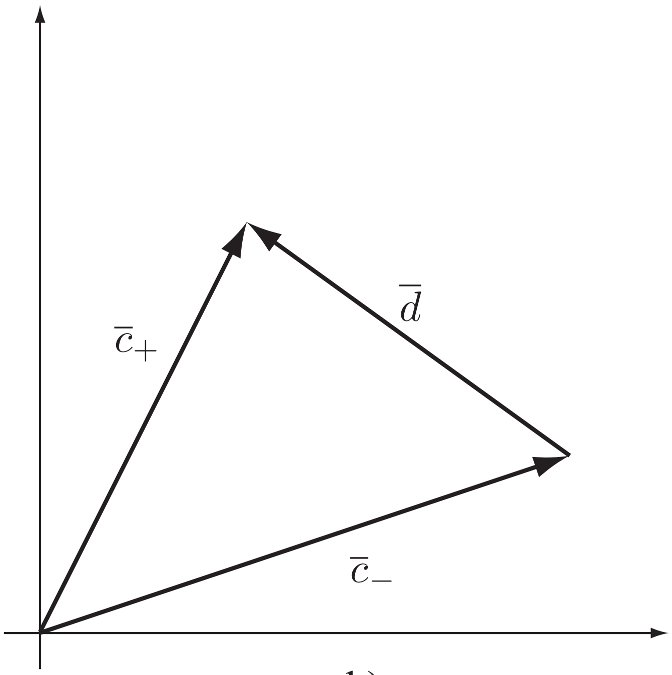
a)

A Simple Learning Algorithm

Step 2.

Next, we construct the vector $\overline{d}$ such that $\overline{c}_+ = \overline{c}_- + \overline{d}$ or,

$$\overline{d} = \overline{c}_+ - \overline{c}_- .$$



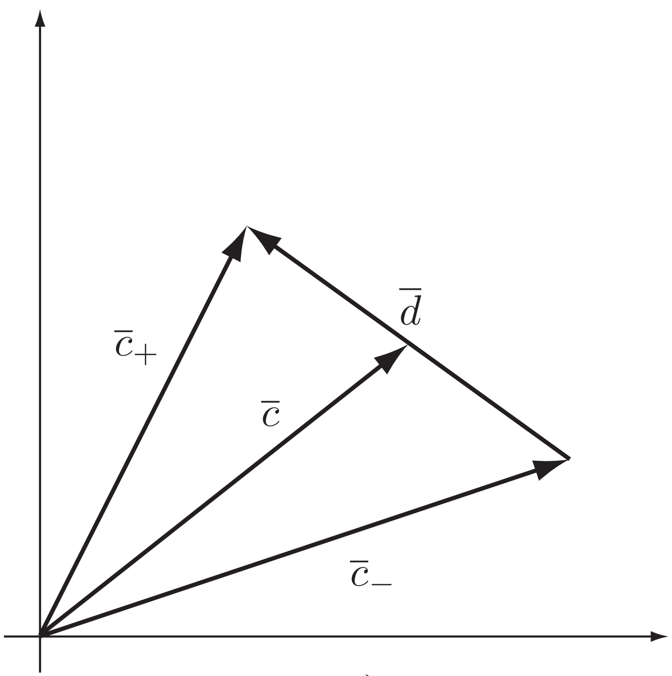
b)

A Simple Learning Algorithm

Step 3.

Compute the midpoint, $\overline{c}$, between the two means $\overline{c}_+$ and $\overline{c}_-$ such that

$$\overline{c} = \frac{1}{2}(\overline{c}_+ + \overline{c}_-).$$

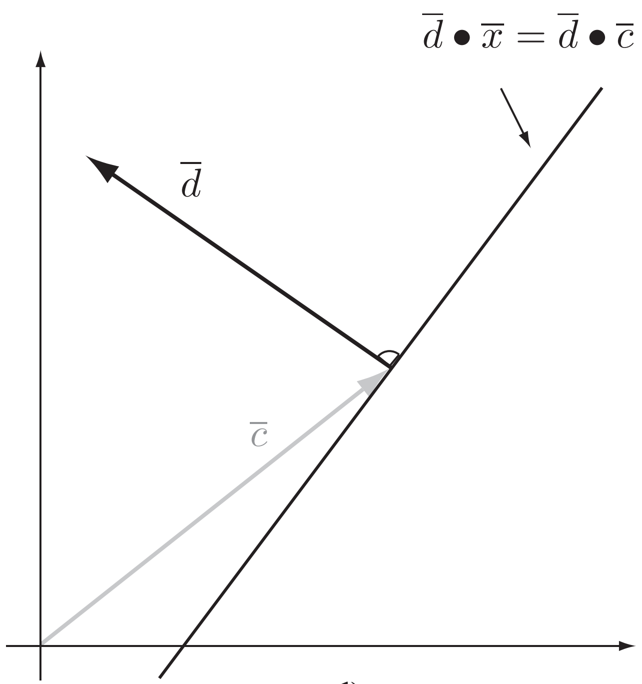


c)

A Simple Learning Algorithm

Step 4.

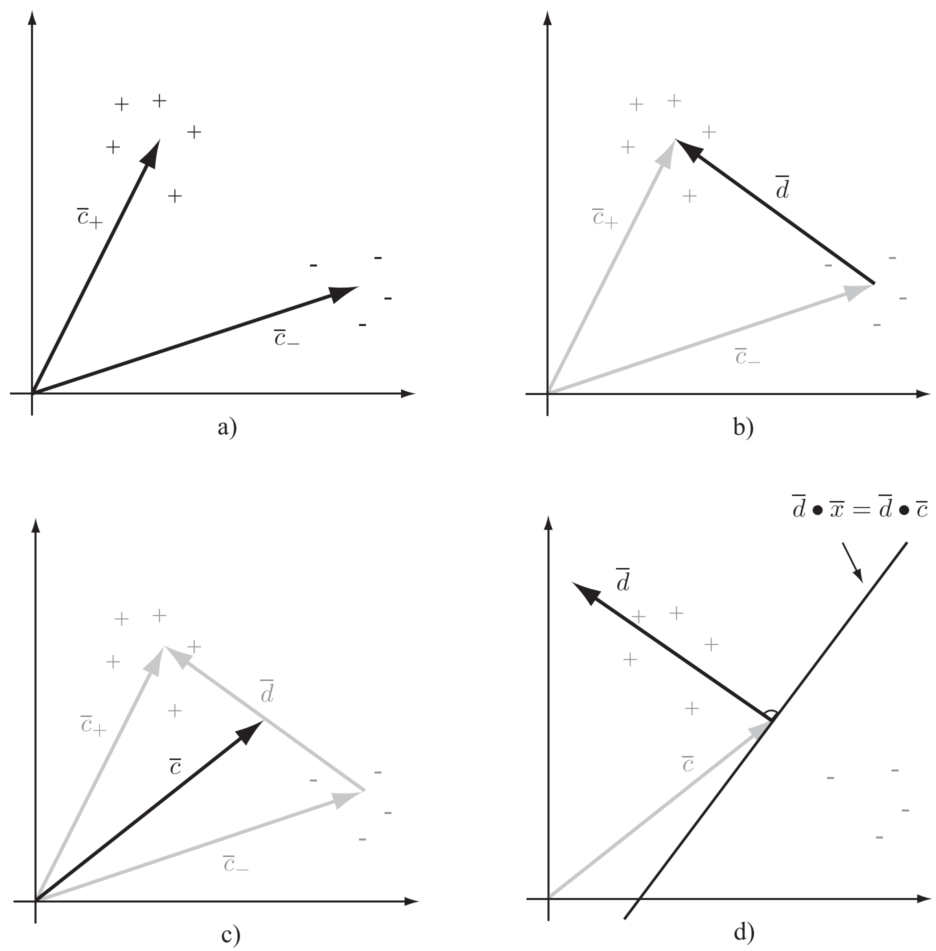
We translate the vector $\overline{d}$ so that it is rooted in the average object $\overline{c}$ and we construct a line perpendicular to $\overline{d}$ through $\overline{c}$. We can now interpret this line as a decision surface $\overline{w} \bullet \overline{x} = b$ with with $\overline{w} = \overline{d}$ and $b = \overline{d} \bullet \overline{c}$, graphically,



d)

A Simple Learning Algorithm

Putting this all together.



A Simple Learning Algorithm

Step 5.

Finally, given our decision surface above we obtain the model,

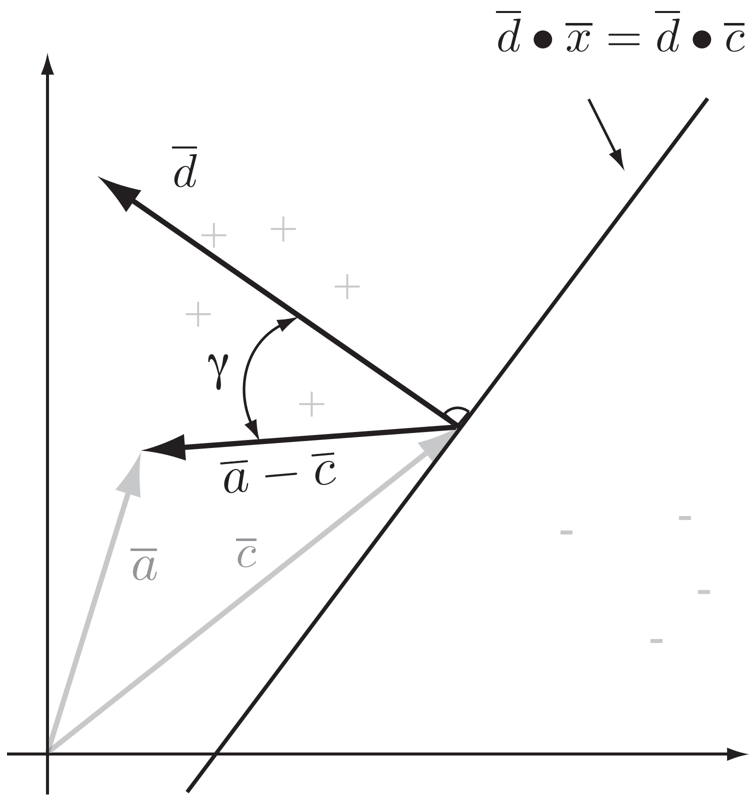
$$
\begin{aligned}
\hat{f}(\overline{x}) &= \operatorname{sign}(\overline{d} \bullet \overline{x} - \overline{d} \bullet \overline{c}) \\
&= \operatorname{sign}\left((\overline{x} - \overline{c}) \bullet \overline{d}\right) \\
&= \operatorname{sign}\left(|\overline{x} - \overline{c}||\overline{d}| \cos \gamma\right),
\end{aligned}
$$

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

A Simple Learning Algorithm

We can illustrate this with a point $\overline{a}$,

$$\hat{f}(\overline{a}) = \mathrm{sign}\left(|\overline{a} - \overline{c}||\overline{d}|\cos\gamma\right),$$



$$\hat{f}(\overline{a}) = \begin{cases} +1 & \text{if } \gamma \leq 90^\circ \\ -1 & \text{if } \gamma > 90^\circ \end{cases}$$

A Simple Learning Algorithm

We can derive an algebraic form of our model from the definitions of $\overline{c}$ and $\overline{d}$,

$$\begin{aligned}\hat{f}(\overline{x}) &= \text{sign}\left((\overline{x} - \overline{c}) \bullet \overline{d}\right) \\ &= \text{sign}\left(\left[\overline{x} - \frac{1}{2}(\overline{c}_{+} + \overline{c}_{-})\right] \bullet (\overline{c}_{+} - \overline{c}_{-})\right).\end{aligned}$$

This shows that the model uses the class means in order to classify unknown points.

A Simple Learning Algorithm

Limitations.

Outliers can distort the orientation of the decision surface which leads to misclassification errors.