Artificial Neural Networks

Lecture 4: Cover's Theorem

Dr Juan Neirotti

j.p.neirotti@aston.ac.uk

2023

Why do we need to go beyond Perceptrons?

- Perceptrons are simple networks (with only one unit) that are easy to analyze.
- Is it possible to find $\boldsymbol{w} \in \Omega \subset \mathbb{R}^d$ and $w_0 \in \mathbb{R}$ for any $\mathcal{D} = \{(\boldsymbol{x}_\ell, t_\ell)\}_{\ell=1}^{L}$ with $\boldsymbol{x}_\ell \in \mathscr{X} \subset \mathbb{R}^d$ and $t_\ell \in \{\pm 1\}$ such that $t_\ell \mathrm{sgn}\left(\langle \boldsymbol{w} \,|\boldsymbol{x}_\ell \rangle + w_0\right) > 0$?
- No, it is not. Can we measure how many classifications can be solved by a perceptron?

Linearly Separable Dichotomies

- Let $\mathscr{X} \subset \mathbb{R}^d$ be an arbitrary set of vectors. A dichotomy $\{\mathscr{X}^+, \mathscr{X}^-\}$ of $\mathscr{X}$ is linearly separable if and only there is a vector $\boldsymbol{w} \in \mathbb{R}^d$ and a scalar $w_0 \in \mathbb{R}$ such that

$$\begin{aligned} \langle \boldsymbol{w} \,|\boldsymbol{x}\,\rangle + w_0 > 0 & \qquad \forall \boldsymbol{x} \in \mathscr{X}_d^+ \\ \langle \boldsymbol{w} \,|\boldsymbol{x}\,\rangle + w_0 < 0 & \qquad \forall \boldsymbol{x} \in \mathscr{X}_d^- . \end{aligned}$$

- The dichotomy is homogeneous if it is linearly separable and $w_0 = 0$.

Linearly Separable Dichotomies

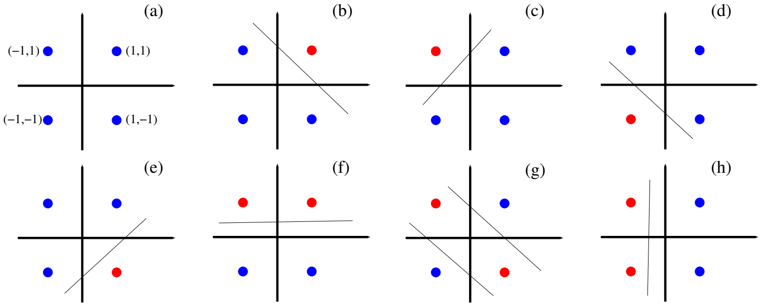

Figure: Half of all possible dichotomies of vectors $\boldsymbol{x} \in \{\pm 1\}^2$. Observe that (g) is not linearly separable.

Augmented Homogeneous Dichotomies

- Suppose that the dichotomy $\{\mathscr{X}^+, \mathscr{X}^-\}$ of $\mathscr{X} \subset \mathbb{R}^d$ is linearly separable with for $\boldsymbol{w} \in \mathbb{R}^d$ $\mathbb{R} \ni w_0 \neq 0$. Then, for all $\boldsymbol{x} \in \mathscr{X}$ we define $\boldsymbol{y} \in \{1\} \times \mathscr{X} = \mathscr{Y}$ such that $|\boldsymbol{y}\rangle = |1\rangle \otimes |\boldsymbol{x}\rangle = |1, \boldsymbol{x}\rangle$ and $|\boldsymbol{W}\rangle = |w_0, \boldsymbol{w}\rangle$ where $(w_0, \boldsymbol{w}) \in \mathbb{R}^{d+1}$. Then, the dichotomy $\{\mathscr{Y}^+, \mathscr{Y}^-\}$ of $\mathscr{Y} \subset \mathbb{R}^{d+1}$ is homogeneous for $\boldsymbol{W} \in \mathbb{R}^{d+1}$.
- Observe that the linear dependencies of the sets $\mathscr{Y}$, $\mathscr{Y}^+$ and $\mathscr{Y}^-$ are identical to the correspondent linear dependencies of $\mathscr{X}$, $\mathscr{X}^+$ and $\mathscr{X}^-$.
- Given this last property we will only consider homogeneous dichotomies in the following discussion.

Augmented Homogeneous Dichotomies

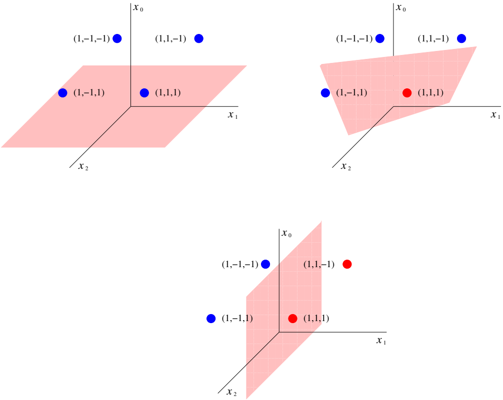

Figure: Augmented dichotomies of vectors $\boldsymbol{x} \in \{\pm 1\}^2$.

Vectors in General Position

- A set $\mathscr{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L\}$ is linearly independent if the equation:
$$\alpha_1 |\boldsymbol{x}_1\rangle + \cdots + \alpha_N |\boldsymbol{x}_N\rangle = |\mathbf{0}\rangle$$
with $\alpha_i \in \mathbb{R}$ admits only the solution $\alpha_i = 0$ for all $1 \leq i \leq N$.
- A set of N vectors is in *general position* in d-dimensional space if every subset of d or fewer vectors is linearly independent. Thus: $\forall\, \mathscr{X}_\ell \subseteq \mathscr{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ with $\boldsymbol{x}_j \in \mathbb{R}^d$ and $|\mathscr{X}_\ell| \leq d$ for all ℓ are linearly independent, then $\mathscr{X}$ is in general position.
- Observe that the set of vertexes of the top square in $\mathbb{R}^3$: $\mathscr{X} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \right\}$ is a set in general position.

Lemma

- Let $\mathscr{X}^+$ and $\mathscr{X}^-$ be subsets of $\mathbb{R}^{d+1}$ and let $\boldsymbol{y} \in \mathbb{R}^{d+1}$ a point other than the origin. Then the dichotomies $\{\mathscr{X}^+ \cup \{\boldsymbol{y}\}, \mathscr{X}^-\}$ and $\{\mathscr{X}^+, \mathscr{X}^- \cup \{\boldsymbol{y}\}\}$ are both homogeneously linearly separable if and only if $\{\mathscr{X}^+, \mathscr{X}^-\}$ is homogeneously linearly separable by a d-dimensional subspace containing $\boldsymbol{y}$.
- Let us define the set $\mathscr{W}$ of separating vectors $\mathscr{W} = \{\mathbb{R}^{d+1} \ni \boldsymbol{w} : \langle \boldsymbol{w} \,|\boldsymbol{x}\,\rangle > 0,\, \boldsymbol{x} \in \mathscr{X}^+; \langle \boldsymbol{w} \,|\boldsymbol{x}\,\rangle < 0,\, \boldsymbol{x} \in \mathscr{X}^-\}$.

Proof

- The dichotomy $\{\mathscr{X}^+ \cup \{\boldsymbol{y}\}, \mathscr{X}^-\}$ is homogeneously linearly separable iff there exists $\boldsymbol{w}_1 \in \mathscr{W}$ such that $\langle \boldsymbol{w}_1 | \boldsymbol{y} \rangle > 0$, $\langle \boldsymbol{w}_1 | \boldsymbol{x}^+ \rangle > 0$ for all $\boldsymbol{x}^+ \in \mathscr{X}^+$ and $\langle \boldsymbol{w}_1 | \boldsymbol{x}^- \rangle < 0$ for all $\boldsymbol{x}^- \in \mathscr{X}^-$; and the dichotomy $\{\mathscr{X}^+, \mathscr{X}^- \cup \{\boldsymbol{y}\}\}$ is homogeneously linearly separable iff there exists $\boldsymbol{w}_2 \in \mathscr{W}$ such that $\langle \boldsymbol{w}_2 | \boldsymbol{y} \rangle < 0$, $\langle \boldsymbol{w}_2 | \boldsymbol{x}^+ \rangle > 0$ for all $\boldsymbol{x}^+ \in \mathscr{X}^+$ and $\langle \boldsymbol{w}_2 | \boldsymbol{x}^- \rangle < 0$ for all $\boldsymbol{x}^- \in \mathscr{X}^-$.
- If $\{\mathscr{X}^+ \cup \{\boldsymbol{y}\}, \mathscr{X}^-\}$ and $\{\mathscr{X}^+, \mathscr{X}^- \cup \{\boldsymbol{y}\}\}$ homogeneously linearly separable by $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ respectively, then $|\boldsymbol{w}^\star\rangle = -\langle \boldsymbol{w}_2 | \boldsymbol{y} \rangle |\boldsymbol{w}_1\rangle + \langle \boldsymbol{w}_1 | \boldsymbol{y} \rangle |\boldsymbol{w}_2\rangle$ separates the dichotomy $\{\mathscr{X}^+, \mathscr{X}^-\}$ by the plane $\{\boldsymbol{x} : \langle \boldsymbol{w}^\star | \boldsymbol{x} \rangle = 0\}$ passing through $\boldsymbol{y}$.
- Conversely, if $\{\mathscr{X}^+, \mathscr{X}^-\}$ is homogeneously linearly separable by a hyperplane containing $\boldsymbol{y}$, then there exists a $\boldsymbol{w}^\star \in \mathscr{W}$ such that $\langle \boldsymbol{w}^\star | \boldsymbol{x}^+ \rangle > 0$, $\langle \boldsymbol{w}^\star | \boldsymbol{x}^- \rangle < 0$ and by definition of the hyperplane $\langle \boldsymbol{w}^\star | \boldsymbol{y} \rangle = 0$.

Extension

- Since $\mathscr{W}$ is open there exists an $\epsilon > 0$ such that $|\boldsymbol{w}^\star\rangle \pm \epsilon\,|\boldsymbol{y}\rangle$ are both in $\mathscr{W}$ and $\{\mathscr{X}^+ \cup \{\boldsymbol{y}\}, \mathscr{X}^-\}$ and $\{\mathscr{X}^+, \mathscr{X}^- \cup \{\boldsymbol{y}\}\}$ are both h.l.s. by $|\boldsymbol{w}^\star\rangle \pm \epsilon\,|\boldsymbol{y}\rangle$ respectively.
- Remember that $\langle \boldsymbol{w}_{1,2}\,|\boldsymbol{x}^+\rangle > 0$, $\langle \boldsymbol{w}_{1,2}\,|\boldsymbol{x}^-\rangle < 0$ for all $\boldsymbol{x}^\pm \in \mathscr{X}^\pm$, then:

$$\begin{aligned}
&\left\langle \boldsymbol{x}^+ \right| \left(|\boldsymbol{w}^\star\rangle + \epsilon_1\,|\boldsymbol{y}\rangle\right) > 0 \qquad && \left\langle \boldsymbol{x}^+ \right| \left(|\boldsymbol{w}^\star\rangle - \epsilon_2\,|\boldsymbol{y}\rangle\right) > 0 \\
&\left\langle \boldsymbol{x}^- \right| \left(|\boldsymbol{w}^\star\rangle + \epsilon_1\,|\boldsymbol{y}\rangle\right) < 0 \qquad && \left\langle \boldsymbol{x}^- \right| \left(|\boldsymbol{w}^\star\rangle - \epsilon_2\,|\boldsymbol{y}\rangle\right) < 0
\end{aligned}$$

Theorem

- There are $C(N, d+1)$ homogeneously linearly separable dichotomies of N points in general position in $\mathbb{R}^{d+1}$,where

$$C(N, d+1) = 2\sum_{k=0}^{d} \binom{N-1}{k}.$$

- $\binom{m}{n} = \frac{m!}{(m-n)!n!}$.
- $d+1$ is the dimension of the augmented representation of the vectors.

Proof

- Let us suppose we have a set $\mathscr{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N : \boldsymbol{x}_\ell \in \mathbb{R}^{d+1}; 1 \leq \ell \leq N\}$, that has a number $C(N, d+1)$ of homogeneously linearly separable dichotomies.
- Consider a new point $\boldsymbol{x}_{N+1} \in \mathbb{R}^{d+1}$ in such a way that $\mathscr{X} \cup \{\boldsymbol{x}_{N+1}\}$ is in general position.

Proof

- If a dichotomy $\{\mathscr{X}^+, \mathscr{X}^-\}$ is separable, by the lemma we have that both $\{\mathscr{X}^+ \cup \{\boldsymbol{x}_{N+1}\}, \mathscr{X}^-\}$ and $\{\mathscr{X}^+, \mathscr{X}^- \cup \{\boldsymbol{x}_{N+1}\}\}$ must be separable if and only if there exists a separating vector $\boldsymbol{w}$ for $\{\mathscr{X}^+, \mathscr{X}^-\}$ lying in the d hyperplane orthogonal to $\boldsymbol{x}_{N+1}$, i.e. $\langle \boldsymbol{w} \,|\boldsymbol{x}_{N+1} \rangle = 0$.
- Given that for each dichotomy $\{\mathscr{X}^+, \mathscr{X}^-\}$ there is at least a vector $\boldsymbol{w}$ such that $\langle \boldsymbol{w} \,|\boldsymbol{x}_{N+1} \rangle = 0$, observe that for all $\boldsymbol{x}_\pm \in \mathscr{X}^\pm$ we have that

$$|\boldsymbol{x}_\pm\rangle = \frac{\langle \boldsymbol{x}_{N+1} \,|\boldsymbol{x}_\pm \rangle}{\|\boldsymbol{x}_{N+1}\|_2} \,|\boldsymbol{x}_{N+1}\rangle + |\boldsymbol{x}_{\pm,\perp}\rangle \,,$$

where $\boldsymbol{x}_{\pm,\perp}$ is projection of $\boldsymbol{x}_\pm$ to the space perpendicular to $\boldsymbol{x}_{N+1}$, then $\langle \boldsymbol{w} \,|\boldsymbol{x}_\pm \rangle = \langle \boldsymbol{w} \,|\boldsymbol{x}_{\pm,\perp} \rangle$.

Proof

- Then a dichotomy $\{\mathscr{X}^+, \mathscr{X}^-\}$ is separable by $\boldsymbol{w}$ with $\langle \boldsymbol{w} | \boldsymbol{x}_{N+1} \rangle = 0$, if and only if the projection of the set $\mathscr{X} = \mathscr{X}^+ \cup \mathscr{X}^-$ onto the d-dimensional orthogonal subspace to $\boldsymbol{x}_{N+1}$ is separable . (This would not be possible in all cases if $\boldsymbol{x}_{N+1}$ were a linear combination of the elements of $\mathscr{X}$, that is why it is required that $\mathscr{X} \cup \{\boldsymbol{x}_{N+1}\}$ is in general position). By hypothesis, a set of N vectors in general position in a d-dimensional space must have a total of $C(N, d)$ dichotomies.

Proof

- Putting these two components together we have that:

$$
\begin{aligned}
C(N+1, d+1) &= C(N, d+1) + C(N, d) \\
&= C(N-1, d+1) + 2C(N-1, d) + C(N-1, d-1) \\
&= C(N-2, d+1) + 3C(N-2, d) + 3C(N-2, d-1) + C(N-2, d-2) \\
&\vdots \\
&= \binom{m}{0} C(N+1-m, d+1) + \binom{m}{1} C(N+1-m, d) + \cdots + \binom{m}{m} C(N+1, d+1-m) \\
&\vdots \\
&= \binom{N}{0} C(1, d+1) + \binom{N}{1} C(1, d) + \cdots + \binom{N}{N} C(1, d+1-N) \\
C(N+1, d+1) &= \sum_{k=0}^{N} \binom{N}{k} C(1, d+1-k).
\end{aligned}
$$

Proof

- Observe that:
$$C(1,j) = \begin{cases} 2 & j \geq 1 \\ 0 & j < 1, \end{cases}$$
thus
$$C(N, d+1) = 2\sum_{k=0}^{d} \binom{N-1}{k}. \clubsuit$$

Total number of binary functions.

- |We have consider the case in $\mathbb{R}^2$ ($d = 2$) with 4 vectors ($N = 4$) $\mathscr{X} = \left\{ \binom{1}{1}, \binom{-1}{1}, \binom{1}{-1}, \binom{-1}{-1} \right\}$, that corresponds to the set of augmented vectors in $\mathbb{R}^3$, that posses 14 dichotomies:

$$C(4,3) = 2\left\{ \binom{3}{0} + \binom{3}{1} + \binom{3}{2} \right\} = 2\{1 + 3 + 3\} = 14.$$

- In the general case we have $C(N, d+1)$ and 2^N as the number of dichotomies and possible classifications respectively.

.

Total number of binary functions.

- Observe that the the fraction of dichotomies decays with the dimension of the space:

$$\frac{C(2^d, d+1)}{2^{2^d}} = \frac{1}{2^{2^d}} \sum_{k=0}^{d} \binom{2^d - 1}{k}$$

d	2^{2^d}	C	%
1	4	4	100
2	16	14	87.5
3	256	104	40.6
4	65,536	1,882	2.87
5	$4.3\ 10^9$	94,572	$2.2\ 10^{-3}$
6	$1.8\ 10^{19}$	$5.03\ 10^6$	$2.7\ 10^{-11}$