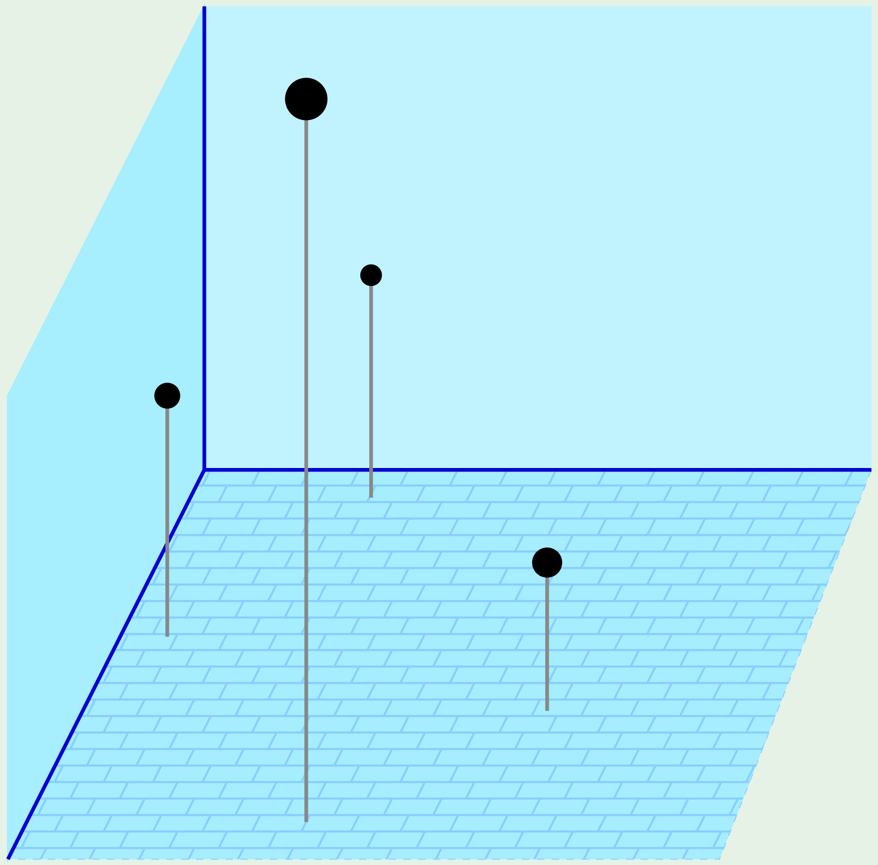
VC dimension of perceptrons

For $d = 2$, $d_{\text{VC}} = 3$

In general, $\boxed{d_{\text{VC}} = d + 1}$



We will prove two directions:

$$d_{\text{VC}} \leq d + 1$$

$$d_{\text{VC}} \geq d + 1$$

Here is one direction

A set of $N = d+1$ points in $\mathbb{R}^d$ shattered by the perceptron:

$$\mathrm{X} = \begin{bmatrix} \text{—}\ \mathbf{x}_1^{\intercal}\ \text{—} \\ \text{—}\ \mathbf{x}_2^{\intercal}\ \text{—} \\ \text{—}\ \mathbf{x}_3^{\intercal}\ \text{—} \\ \vdots \\ \text{—}\ \mathbf{x}_{d+1}^{\intercal}\ \text{—} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & & 0 \\ & \vdots & & \ddots & 0 \\ 1 & 0 & \dots & 0 & 1 \end{bmatrix}$$

X is invertible

Can we shatter this data set?

For any $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{d+1} \end{bmatrix} = \begin{bmatrix} \pm 1 \\ \pm 1 \\ \vdots \\ \pm 1 \end{bmatrix}$, can we find a vector $\mathbf{w}$ satisfying

$$\text{sign}(\mathrm{X}\mathbf{w}) = \mathbf{y}$$

Easy! Just make $\mathrm{X}\mathbf{w} = \mathbf{y}$

which means $\mathbf{w} = \mathrm{X}^{-1}\mathbf{y}$

We can shatter these $d+1$ points

This implies what?

[a] $d_{\mathrm{VC}} = d+1$

[b] $d_{\mathrm{VC}} \geq d+1$ ✓

[c] $d_{\mathrm{VC}} \leq d+1$

[d] No conclusion

Now, to show that $d_{\text{vc}} \leq d + 1$

We need to show that:

[a] There are $d + 1$ points we cannot shatter

[b] There are $d + 2$ points we cannot shatter

[c] We cannot shatter *any* set of $d + 1$ points

[d] We cannot shatter *any* set of $d + 2$ points ✓

Take any $d+2$ points

For any $d+2$ points,

$$\mathbf{x}_1, \cdots, \mathbf{x}_{d+1}, \mathbf{x}_{d+2}$$

More points than dimensions $\implies$ we must have

$$\mathbf{x}_j = \sum_{i \neq j} a_i \, \mathbf{x}_i$$

where not all the a_i's are zeros

So?

$$\mathbf{x}_j = \sum_{i \neq j} a_i \, \mathbf{x}_i$$

Consider the following dichotomy:

$\mathbf{x}_i$'s with non-zero a_i get $\quad y_i = \text{sign}(a_i)$

and $\mathbf{x}_j$ gets $\quad y_j = -1$

No perceptron can implement such dichotomy!

Why?

$$\mathbf{x}_j = \sum_{i \neq j} a_i \, \mathbf{x}_i \quad \Longrightarrow \quad \mathbf{w}^\intercal \mathbf{x}_j = \sum_{i \neq j} a_i \, \mathbf{w}^\intercal \mathbf{x}_i$$

If $y_i = \text{sign}(\mathbf{w}^\intercal \mathbf{x}_i) = \text{sign}(a_i)$, then $a_i \, \mathbf{w}^\intercal \mathbf{x}_i > 0$

This forces $$\mathbf{w}^\intercal \mathbf{x}_j = \sum_{i \neq j} a_i \, \mathbf{w}^\intercal \mathbf{x}_i > 0$$

Therefore, $y_j = \text{sign}(\mathbf{w}^\intercal \mathbf{x}_j) = +1$

Putting it together

We proved $d_{\text{VC}} \leq d+1$ and $d_{\text{VC}} \geq d+1$

$$\boxed{d_{\text{VC}} = d+1}$$

What is $d+1$ in the perceptron?

It is the number of parameters $w_0, w_1, \cdots, w_d$