Appendix: VC dimension – a classic!

The VC dimension is also a way of measuring the complexity of $\mathcal{H}$, but for binary classification.
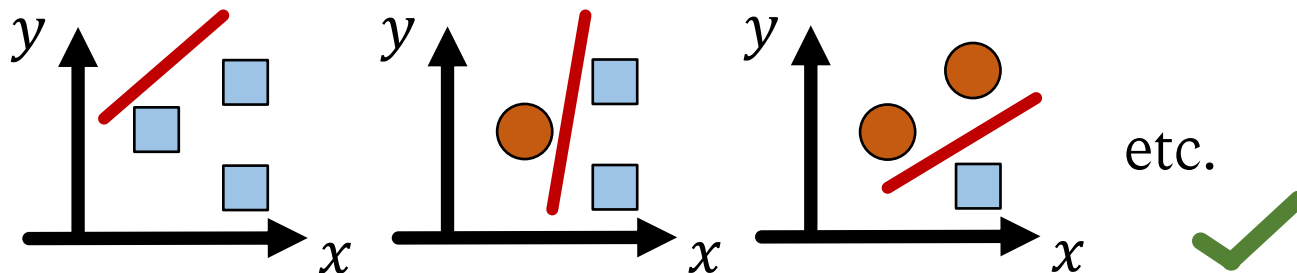
Assume we have a hypothesis set $\mathcal{H}$. We count the number of "equivalent" functions in $\mathcal{H}$ by estimating how many points can be separated by $\mathcal{H}$ at best.

> **VC dimension:** assume a set of points $x = (x_1, x_2, \ldots, x_D)$ and all its possible binary labellings $C = \{(\pm 1)^D\}^{2^D}$. We define the **Vapnik-Chevronenkis** (VC) dimension of $\mathcal{H}$ as the maximum of D where for each $c \in C$, there is an $h_c \in \mathcal{H}$ such that $h_c(x) = c$.
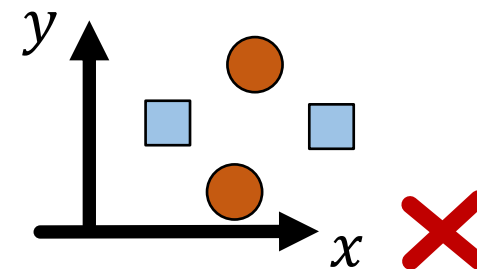
Note: we only have to find one set of points for which this is true!

Example: linear regression

$D = 3$

$D = 4$

etc. ✓

✗

Learning bound: VC dimension

If $\mathcal{H}$ has VC dimension d_{VC}, we have with probability $1 - \delta$ for all h $\in \mathcal{H}$ and for $m \geq d_{VC}$:

$$\left|R(h) - \hat{R}(h)\right| \leq \sqrt{\frac{2d_{VC}\log(em/k)}{m}} + \sqrt{\frac{\log(1/\delta)}{2m}}$$

Examples:

- Linear regression: $d_{VC} = d + 1, d = \#$ features
- ReLU ANN with L layers and n parameters: $d_{VC} \leq C \cdot (nL \log n + nL^2)$
- Sine function: $d_{VC} = \infty$
- K-nearest neighbours: $d_{VC} = \infty$

$C > 0$, independent of n and L