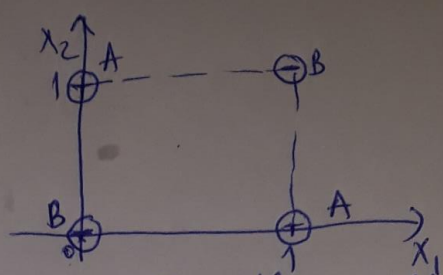
**EXAMPLE 2** XOR is not linearly separable

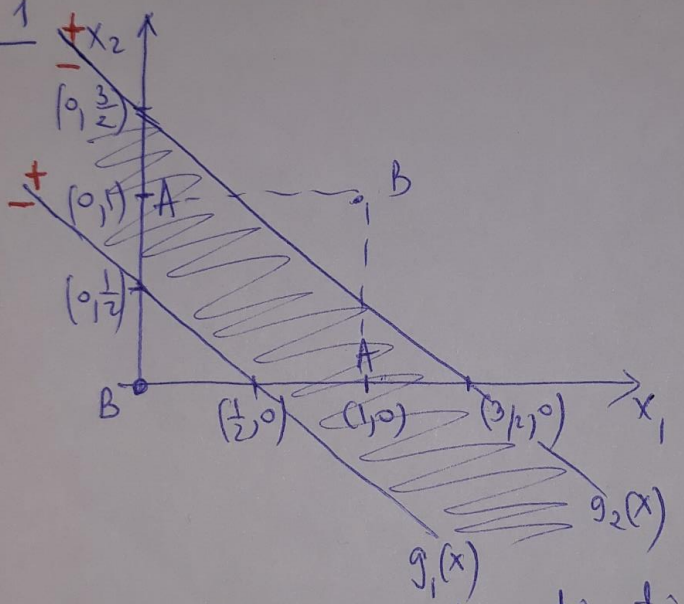
| $X_1$ | $X_2$ | XOR |
|---|---|---|
| 0 | 0 | 0 (B) − |
| 0 | 1 | 1 (A) + |
| 1 | 0 | 1 (A) + |
| 1 | 1 | 0 (B) − |




? How to use linear classifiers to solve this problem?
Idea: use a logical combination of linear classifiers

## The 2-layer perceptron architecture

Layer 1




$$g_1(x) = x_1 + x_2 - \frac{1}{2}$$
$$g_2(x) = x_1 + x_2 - \frac{3}{2}$$

Let $f(x)$ be the "signum" activation function

$$f(x) = \begin{cases} +1 & x > 0 \\ -1 & \text{otherwise} \end{cases}$$

We observe that the classification problem is solved by $f(y_1 - y_2 - 1)$ where $y_1 = f(g_1(x))$

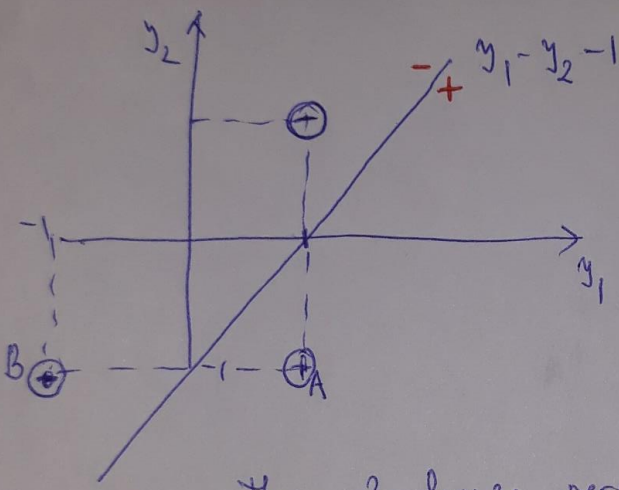
$$y_2 = f(g_2(x))$$

- This calculation can be implemented sequentially
  1. compute $y_1$ and $y_2$ from $x_1$ and $x_2$
  2. compute the decision from $y_1$ and $y_2$

- Each layer in the sequence consists of one or more linear classifiers
- This is therefore a 2-layer perceptron

| Layer 1 | | | | Layer 2 |
|---|---|---|---|---|
| $x_1$ | $x_2$ | $y_1$ | $y_2$ | |
| 0 | 0 | -1 | -1 | B (-) |
| 0 | 1 | 1 | -1 | A (+) |
| 1 | 0 | 1 | -1 | A (+) |
| 1 | 1 | 1 | 1 | B (-) |

The first layer performs a nonlinear mapping that makes the data linearly separable.




=> The 2-layer perceptron