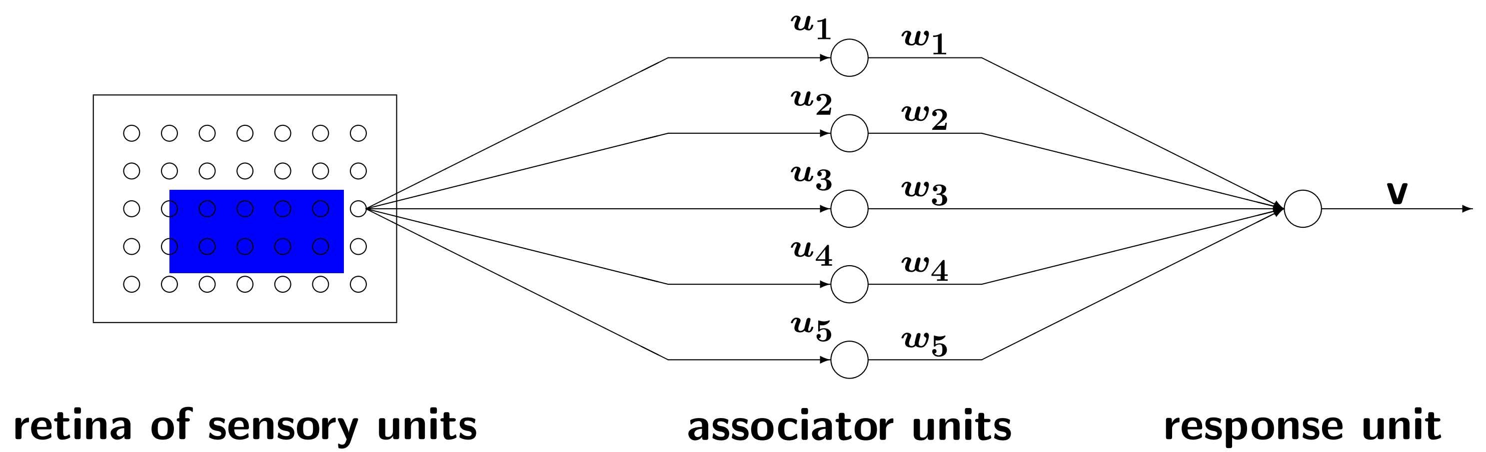
Simple Perceptrons

- (Rosenblatt, 1962): device to learn from experience;
  in particular, to learn to recognize and classify patterns.

- Hertz, Krogh, Palmer: Introduction to the Theory of Neural Computation. Addison-Wesley Publishing Company (1991).

- Simple perceptrons



- Pattern is represented by $(i_1, \ldots, i_m) \in \mathbb{R}^m$. Here, $m = 5$.

- Pattern recognition: $f : \mathbb{R}^m \longrightarrow \{1, -1\}$.

Computation of a Simple Perceptron

- **Let the response unit be a linear threshold unit with threshold 0:**

$$\boldsymbol{v} = \mathrm{sgn}(\sum_{j=1}^{m} \boldsymbol{w}_j \boldsymbol{i}_j) = \mathrm{sgn}(\overline{\boldsymbol{w}} \cdot \bar{\boldsymbol{i}}).$$

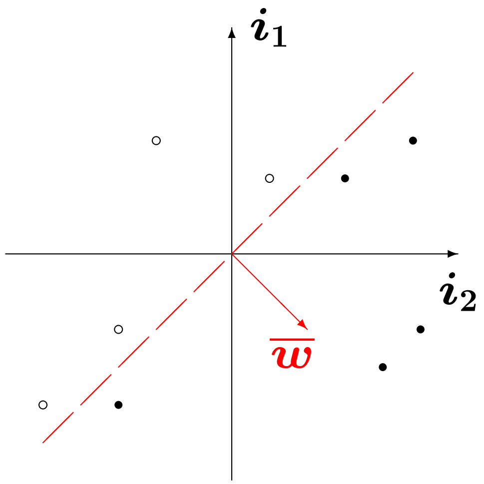
- **A classification problem is a set $\{(\bar{\boldsymbol{i}}^k, \boldsymbol{a}^k) \mid 1 \le \boldsymbol{k} \le \boldsymbol{d}\}$ of input-output (pattern-classification) examples.**

- **Ideally we should find for all $1 \le \boldsymbol{k} \le \boldsymbol{d}$ that $\boldsymbol{v}^k = \mathrm{sgn}(\overline{\boldsymbol{w}} \cdot \bar{\boldsymbol{i}}^k) = \boldsymbol{a}^k$.**

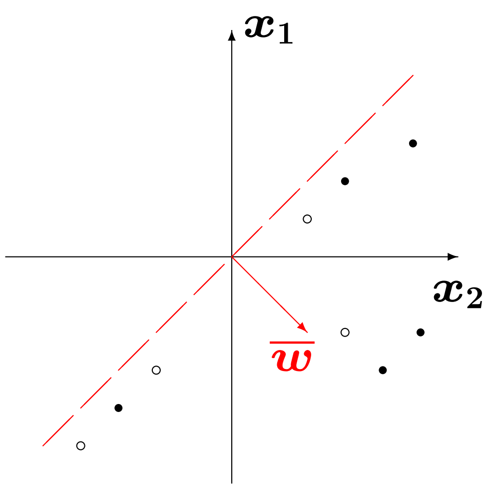
- **In this case, there exists a $\boldsymbol{m} - 1$ dimensional hyperplane through the origin and perpendicular to $\overline{\boldsymbol{w}}$ such that it divides the positive from the negative examples.**

Example

- $\mathbb{R}^2$:

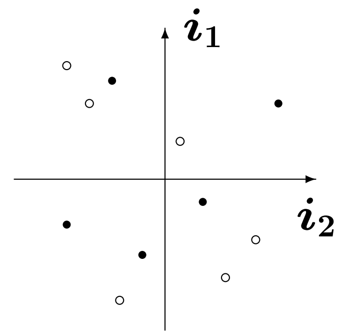


- Let $\overline{x}^k = a^k \overline{i}$:



Linear Separability

- With $\overline{x}^k = a^k \overline{i}^k$ we should find for all $1 \leq k \leq d$ that $\overline{w} \cdot \overline{x}^k > 0$.
- In this case, all examples are on one side of the hyperplane.
- Such a hyperplane does not always exist!



- A classification problem is linearly separable if there exists a hyperplane separating the positive from the negative examples.

Learning

- **Task**: Can we learn $\overline{w}$ such that the perceptron solves the classification problem?
- **Learning algorithm**: Select $\overline{w}$ arbitrarily.
  Do while some examples are misclassified:
  - Present some $\overline{x}^k$ as input to the perceptron (fairly!).
  - Compute v^k.
  - Let $\Delta\overline{w} = \eta(1 - v^k)\overline{x}^k$, where $\eta > 0$ is the learning rate.
  - Update $\overline{w} := \overline{w} + \Delta\overline{w}$.
- Because the problem is linearly separable we find $\overline{w}^*$ such that for all $1 \leq k \leq d$:

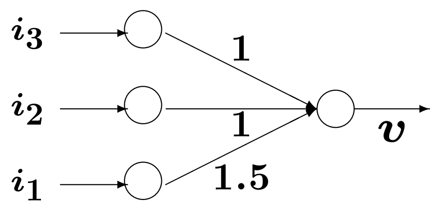
$$\overline{w}^* \cdot \overline{x}^k > 0.$$

- Let $\overline{w}$ be the current weight vector.
- If $\overline{x}^k$ is misclassified, then $\overline{w} \cdot \overline{x}^k \leq 0$.
- In this case:

$$(\overline{w} + \Delta\overline{w}) \cdot \overline{x}^k = \overline{w} \cdot \overline{x}^k + \Delta\overline{w} \cdot \overline{x}^k = \overline{w} \cdot \overline{x}^k + 2\eta\overline{x}^k \cdot \overline{x}^k > \overline{w} \cdot \overline{x}^k.$$

Convergence Theorem

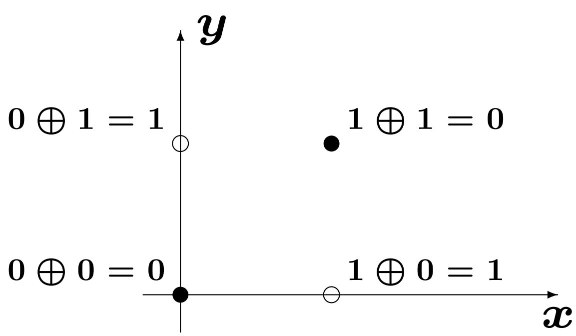
- **Theorem**: **A simple perceptron solves each linearly separable classification problem after a finite number of training steps.**
- **The number of steps is not known in advance, because we do not know $\overline{w}^*$!**
- **Exercise**: **Consider the following simple perceptron:**



  - **What is computed by the perceptron if 1 and -1 are the possible values for i_1 and i_2, whereas i_3 always has value -1?**
  - **Is this problem linearly separable?**
  - **Now suppose that all weights have value 0. What happens if we apply the learning algorithm with $\eta = 0.15$?**

The Minsky-Papert Theory

- Minsky, Papert: Perceptrons. MIT Press: 1972
- How is the structure of the network and the complexity of the units related to learnability?
- Under what conditions can a logical threshold unit learn a classification task?
  - There are formal bounds.
- Consider addition modulo 2 ($\oplus$):



- Many more examples.

Exercises

- Construct a McCulloch-Pitts network which computes the addition modulo 2.
- Consider the classification problem

  $$\{(000,0),\ (001,1),\ (010,1),\ (011,0),\ (100,1),\ (101,0),\ (110,0),\ (111,1)\}$$

  and show that there is no simple perceptron solving it.
- Find a transformation mapping addition modulo 2 onto a classification problem $\mathcal{C}$ in $\mathbb{R}^3$ such that $\mathcal{C}$ is linearly separable. Construct a simple perceptron and train it until it solves $\mathcal{C}$.