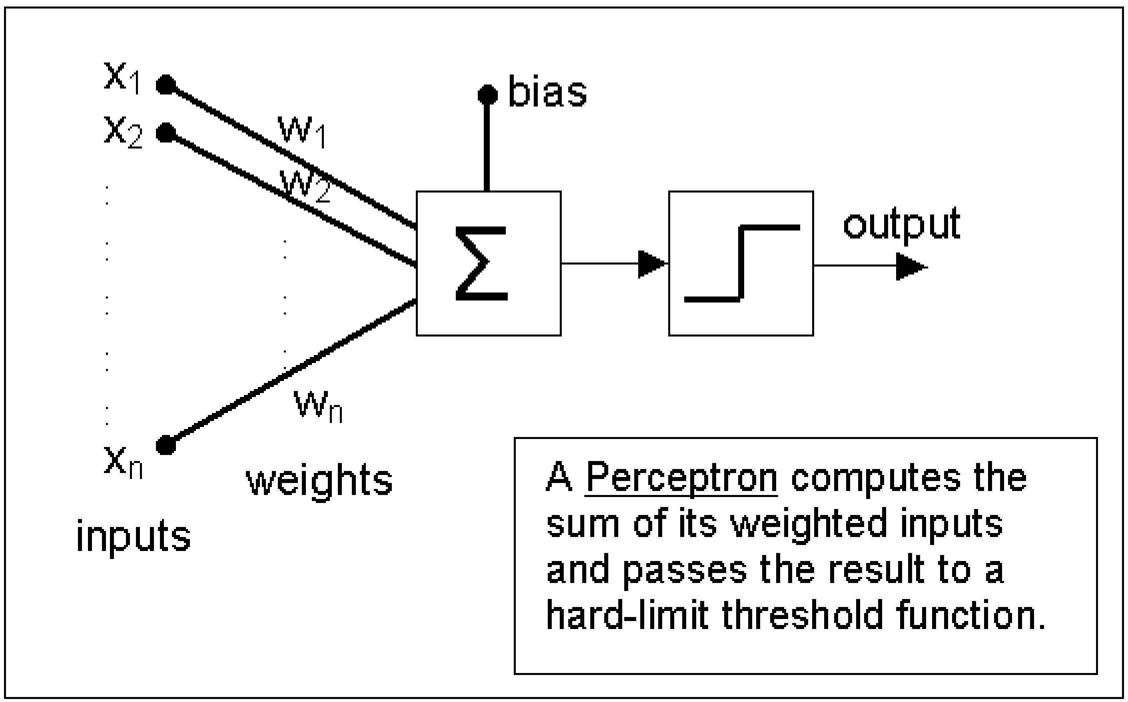
Machine Learning

Lecture 2

Perceptron

Perceptron - Basic

- Perceptron is a type of artificial neural network (ANN)

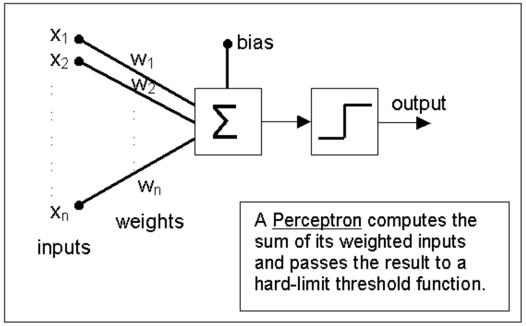


Perceptron - Operation

- It takes a vector of real-valued inputs, calculates a linear combination of these inputs, then output 1 if the result is greater than some threshold and -1 otherwise

$$R = w_0 + w_1x_1 + w_2x_2, \cdots, w_nx_n = w_0 + \sum_{i=1}^{n} w_ix_i$$

$$Y = sign(R) = \begin{cases} +1; & \text{if} \quad R > 0 \\ -1, & \text{otherwise} \end{cases}$$



Perceptron – Decision Surface

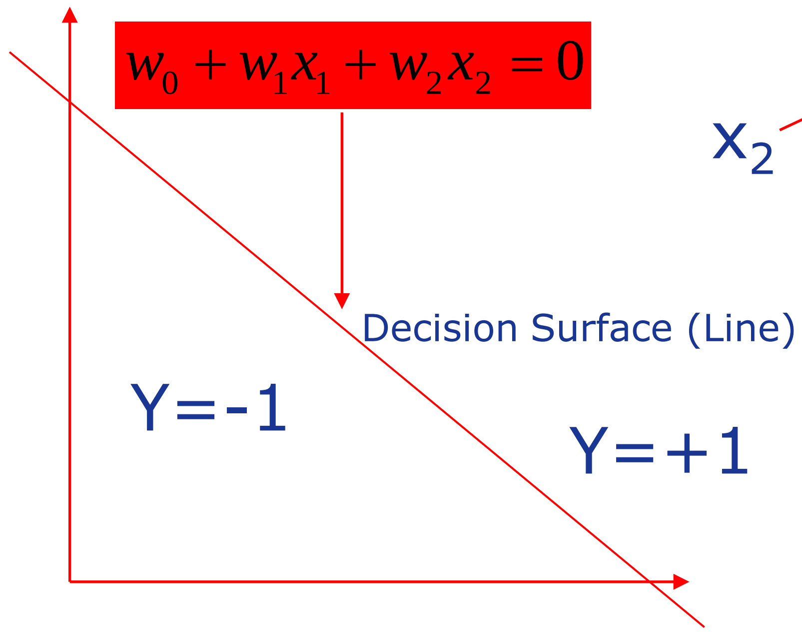
- Perceptron can be regarded as representing a hyperplane decision surface in the n-dimensional **feature space** of instances.

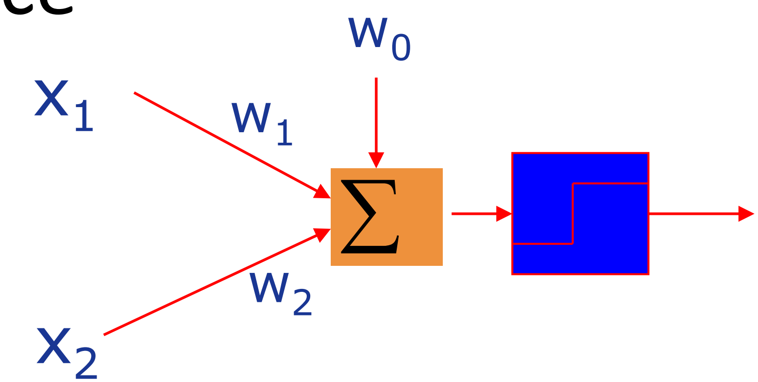
- The perceptron outputs a 1 for instances lying on one side of the hyperplane and a -1 for instances lying on the other side.

- This hyperplane is called the **Decision Surface**

Perceptron – Decision Surface

- In 2-dimensional space

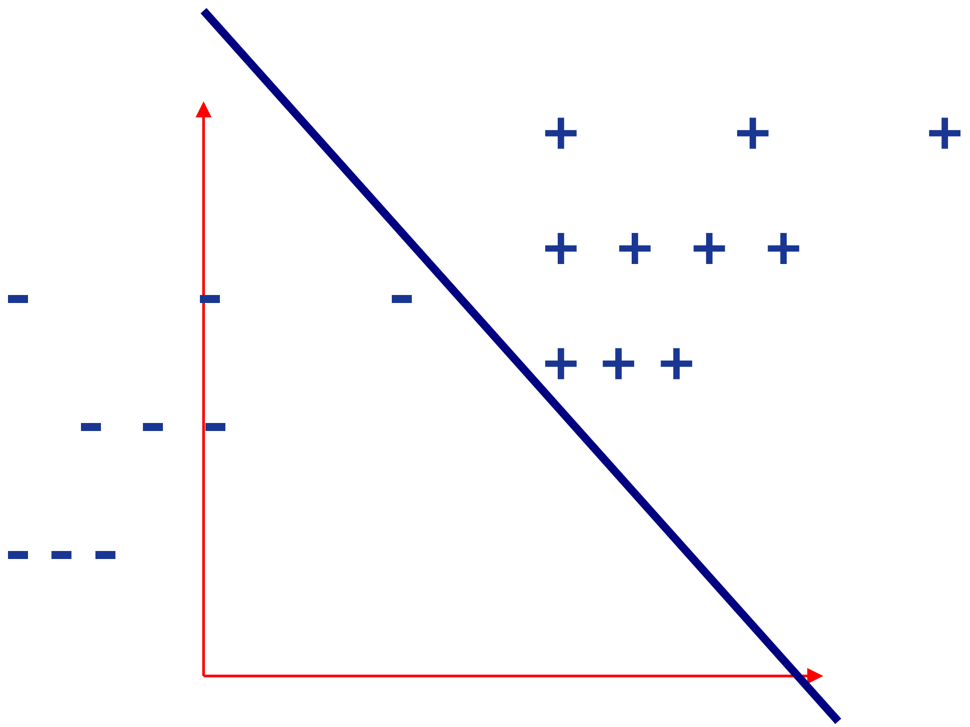




Perceptron – Representation Power

- The Decision Surface is linear

- Perceptron can only solve **Linearly Separable Problems**

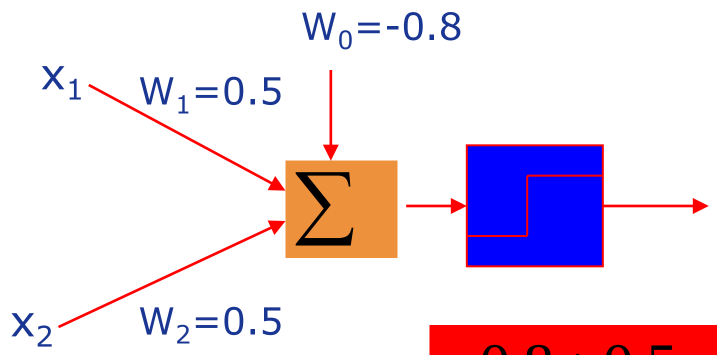


Perceptron – Representation Power

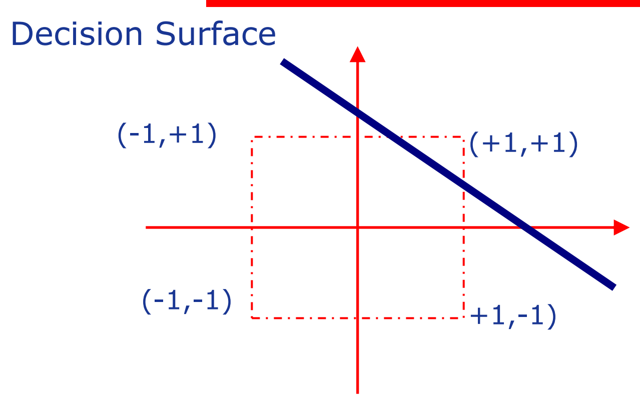
- Can represent many boolean functions: Assume boolean values of 1 (true) and -1 (false)

AND

x1	x2	D
-1	-1	-1
-1	+1	-1
+1	-1	-1
+1	+1	+1

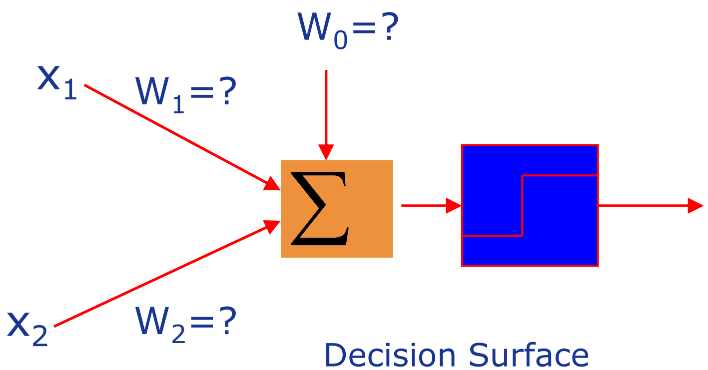


$$-0.8 + 0.5x_1 + 0.5x_2 = 0$$



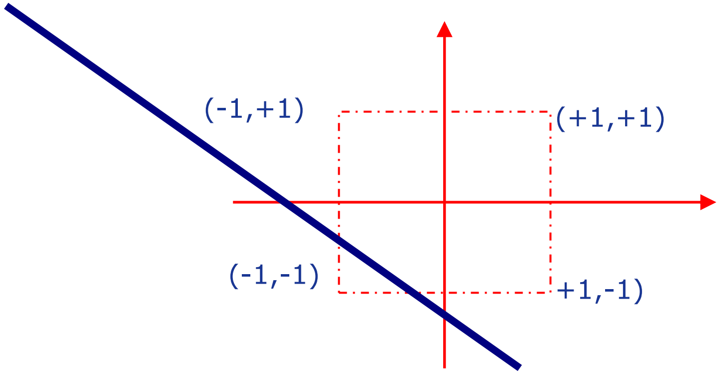
Perceptron – Representation Power

- Can represent many boolean functions: Assume boolean values of 1 (true) and -1 (false)

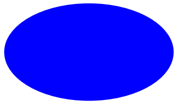


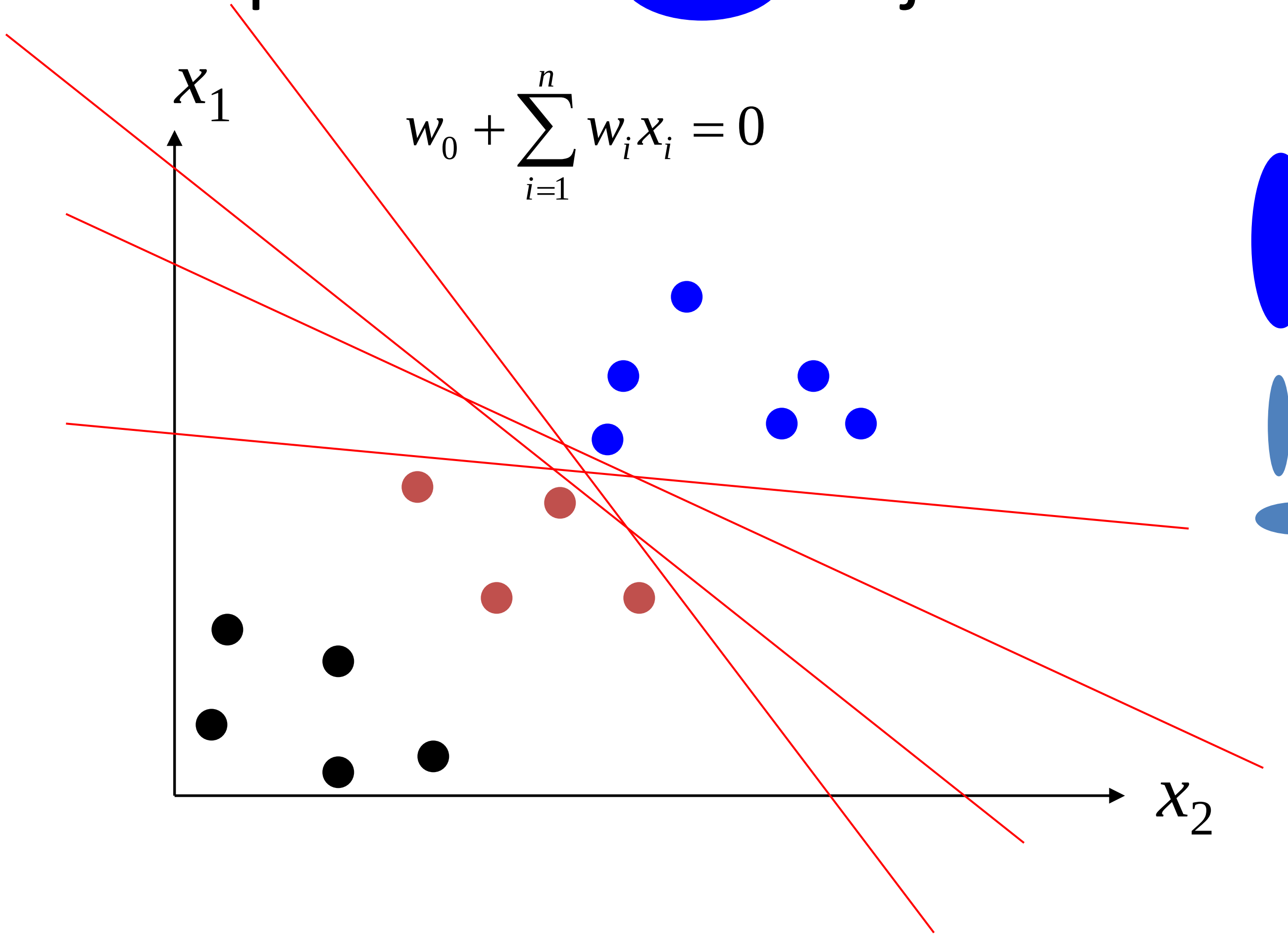
OR

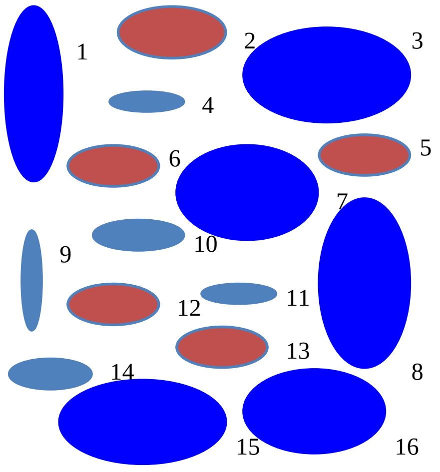
x1	x2	D
-1	-1	-1
-1	+1	+1
+1	-1	+1
+1	+1	+1



Perceptron – Representation Power

- Separate the  objects from the rest





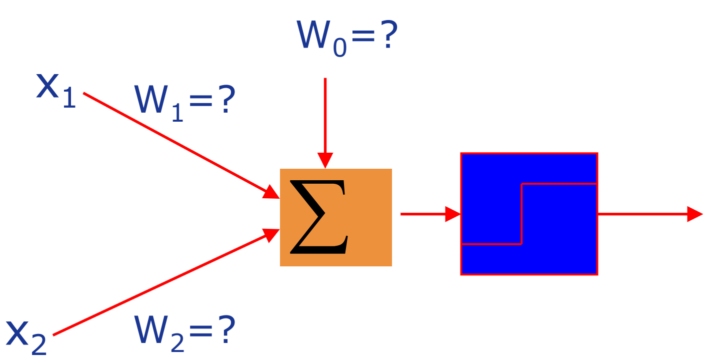
Elliptical blobs (objects)

Perceptron – Representation Power

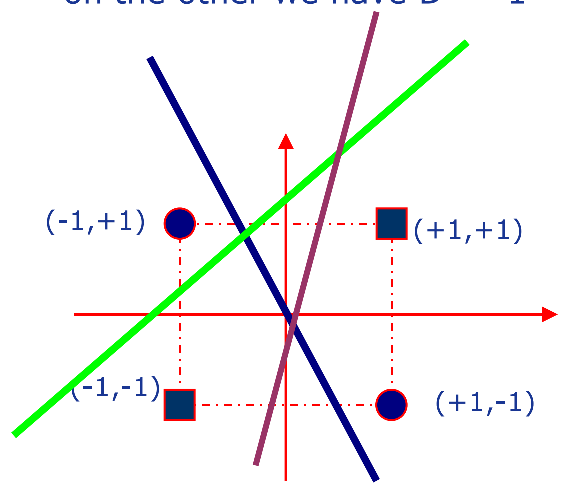
- Some problems are linearly non-separable

XOR

x1	x2	D
-1	-1	-1
-1	+1	+1
+1	-1	+1
+1	+1	-1



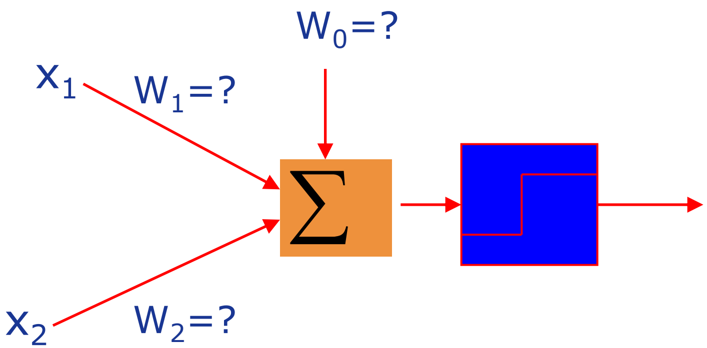
Decision Surface:
It doesn't matter where you place the line (decision surface), it is impossible to separate the space such that on one side we have D = 1 and on the other we have D = -1

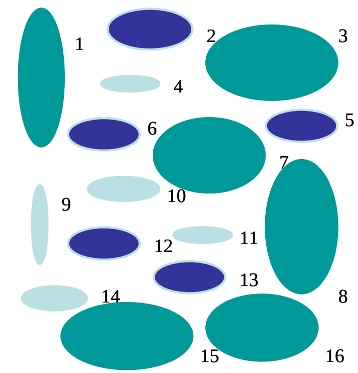


Perceptron Cannot Solve such Problem!

Perceptron – Training Algorithm

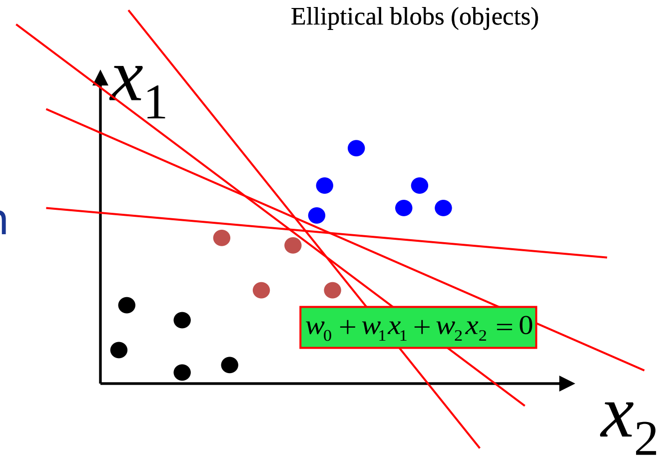
- Separate the  objects from the rest





Elliptical blobs (objects)

We are given the training sample (experience) pairs (X, D), how can we determine the weights that will produce the correct +1 and -1 outputs for the given training samples?



Perceptron – Training Algorithm

- Training sample pairs (X, d), where X is the input vector, d is the input vector's classification (+1 or -1) is iteratively presented to the network for training, *one at a time*, until the process converges

Perceptron – Training Algorithm

- The Procedure is as follows

1. Set the weights to small random values, e.g., in the range (-1, 1)

2. Present X, and calculate

$$R = w_0 + \sum_{i=1}^{n} w_i x_i \qquad Y = sign(R) = \begin{cases} +1; & \text{if} \quad \mathrm{R} > 0 \\ -1, & \text{otherwise} \end{cases}$$

3. Update the weights

$$w_i \leftarrow w_i + \eta (d - y) x_i, i = 1, 2, \cdots, n$$

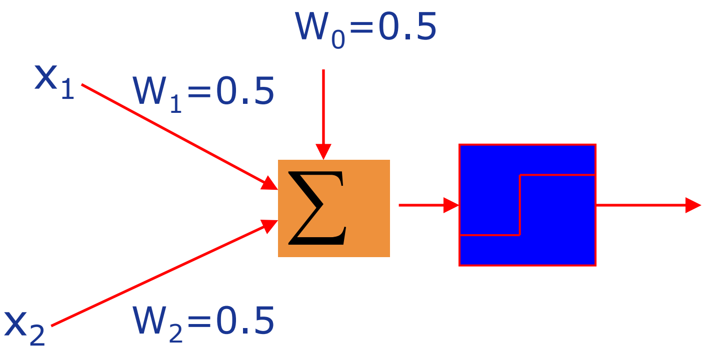
$0 < \eta < 1$ *is the training rate* $x_0 = 1$ (constant)

4. Repeat by going to step 2

Perceptron – Training Algorithm

- Example

x1	x2	D
-1	-1	-1
-1	+1	+1
+1	-1	+1
+1	+1	+1



$$w_i \leftarrow w_i + \eta(d - y)x_i, i = 1,2,\cdots,n$$

Perceptron – Training Algorithm

- Convergence Theorem
  - The perceptron training rule will converge (finding a weight vector correctly classifies all training samples) within a finite number of iterations, **provided the training examples are linearly separable** and provided a sufficiently small η is used.

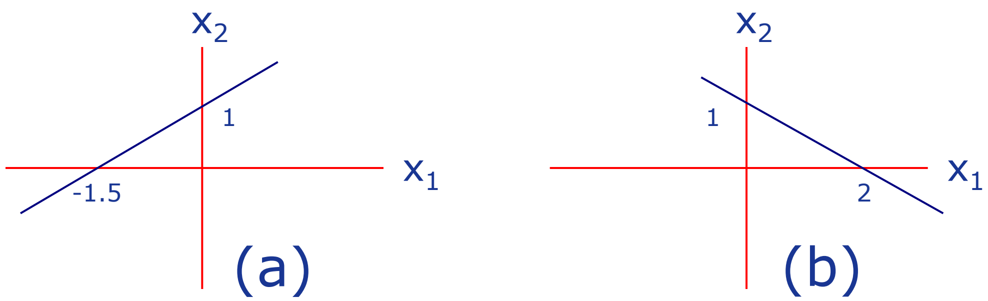
Further Reading

- T. M. Mitchell, Machine Learning, McGraw-Hill International Edition, 1997

Chapter 4

Tutorial/Exercise Questions

1. What is the weight values of a perceptron having the following decision surfaces



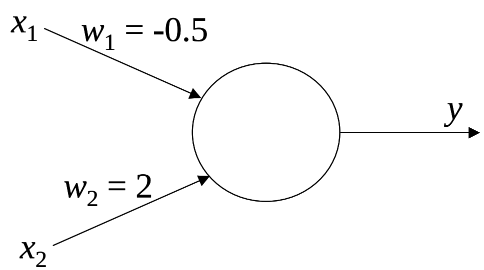
2. Design two-input perceptrons for implementing the following boolean functions

   AND, OR, NAND, NOR

3. A single layer perceptron is incapable of learning simple functions such as XOR (exclusive OR). Explain why this is the case (hint: use the decision boundary)

Tutorial/Exercise Questions

4. A single layer Perceptron is as follows



a) Write down and plot the equation of the decision boundary of this device

b) Change the values of w1 and w2 so that the Perceptron can separate following two-class patterns

Class 1 Patterns: (1, 2), (1.5. 2.5), (1, 3)

Class 2 Patterns: (2, 1.5), (2, 1)