CS559 lecture 9
FDA and Perceptron

Readings: Bishop: Chapter 4.1, Chapter 7.

Outline

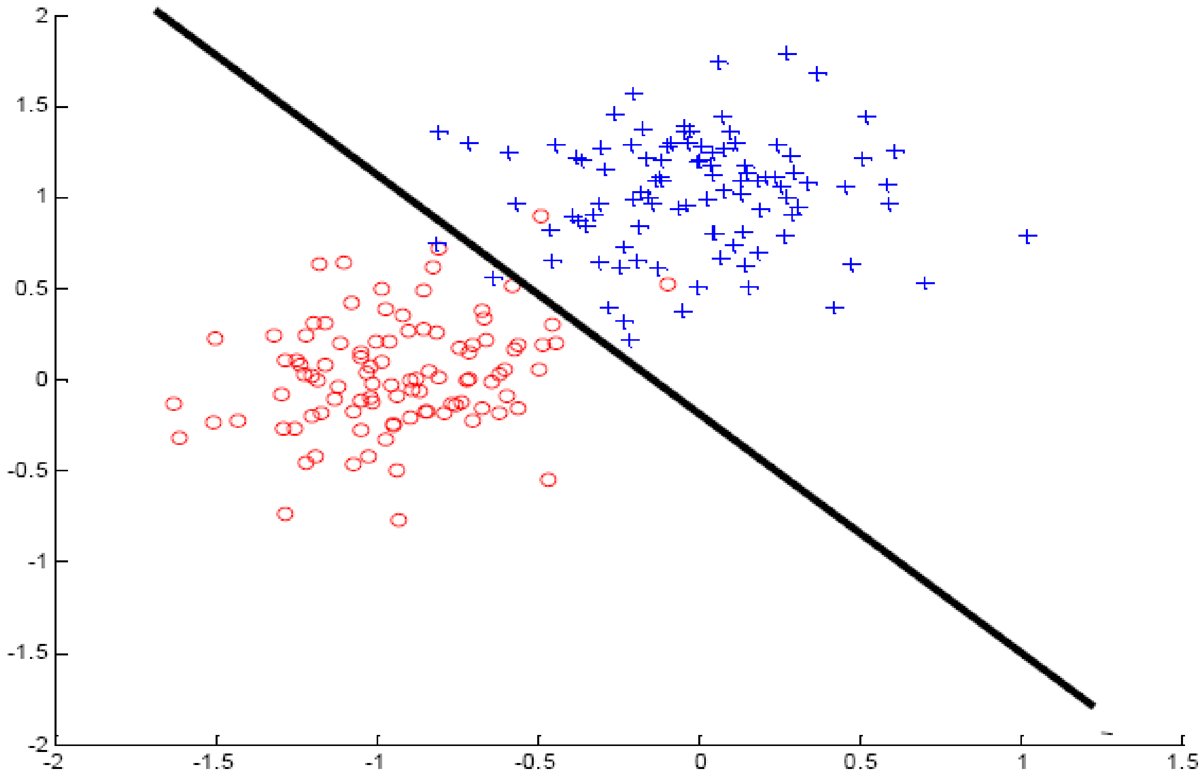
Outline:

- Fisher Linear Discriminant
- Algorithms for linear decision boundary
- **Support vector machines**
- Maximum margin hyperplane.
- Support vectors.
- Support vector machines.

- Extensions to the non-separable case.
- Kernel functions.

Linear decision boundaries

- What models define linear decision boundaries?



Logistic regression model

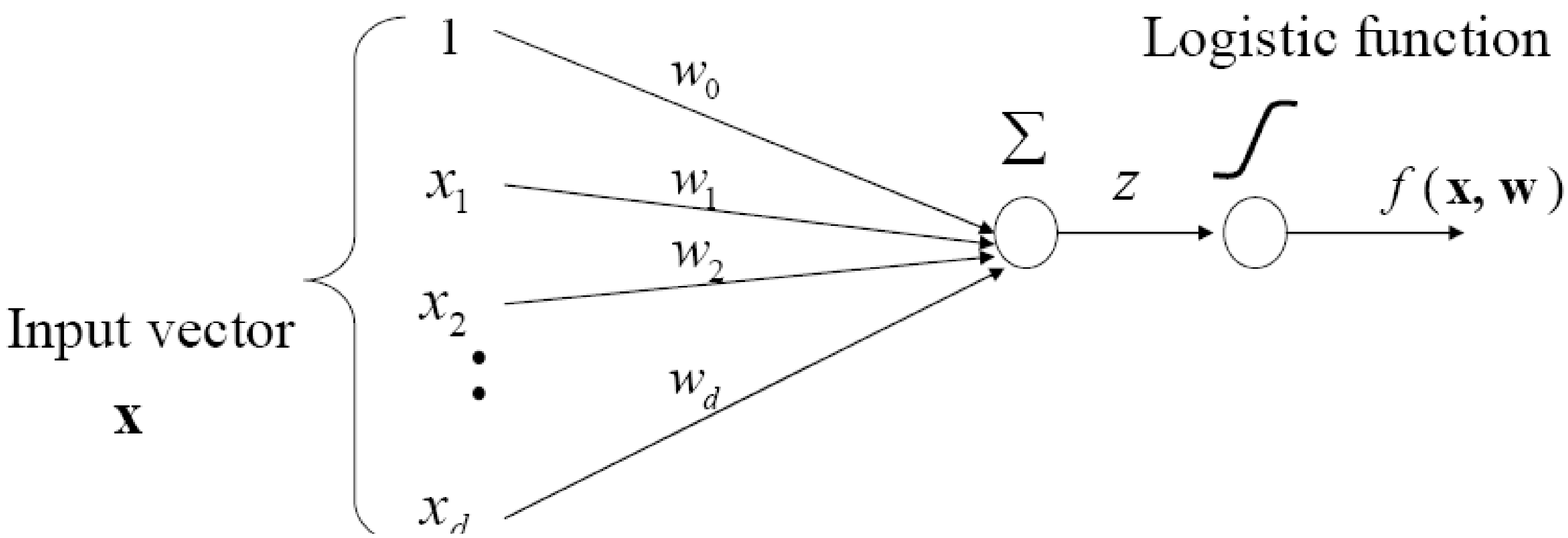
- **Discriminant functions:**

$$g_1(\mathbf{x}) = g(\mathbf{w}^T\mathbf{x}) \qquad\qquad g_0(\mathbf{x}) = 1 - g(\mathbf{w}^T\mathbf{x})$$

- **where**

$g(z) = 1/(1+e^{-z})$ - is a logistic function

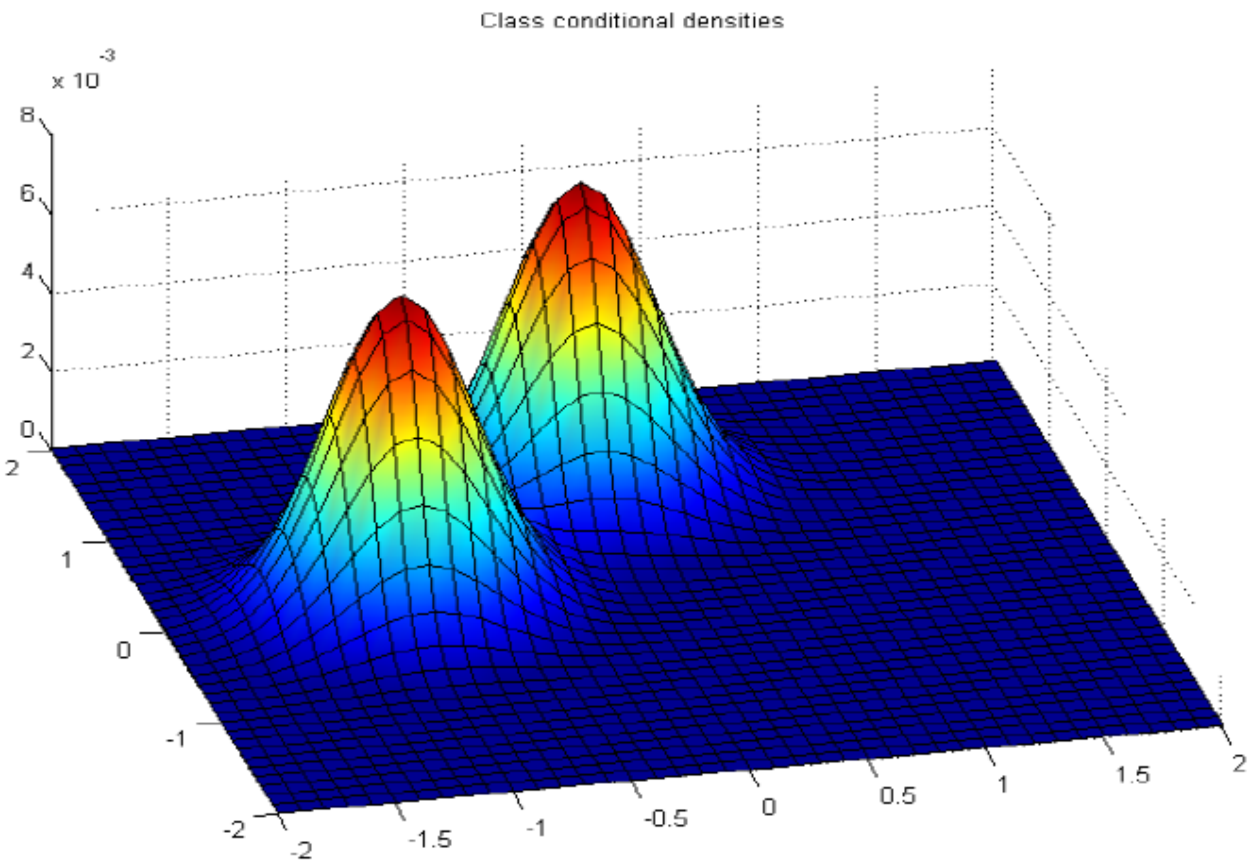
$$f(\mathbf{x},\mathbf{w}) = g_1(\mathbf{w}^T\mathbf{x}) = g(\mathbf{w}^T\mathbf{x})$$



Linear discriminant analysis (LDA)

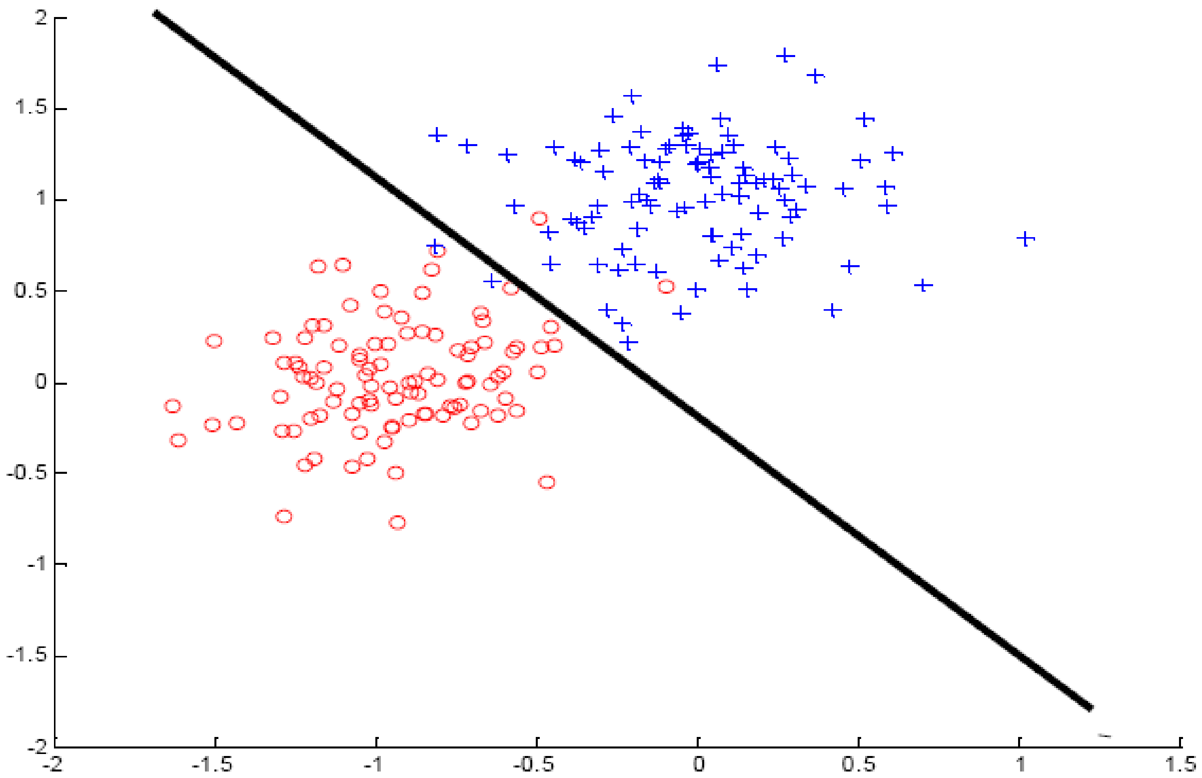
- When covariances are the same

$$\mathbf{x} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}),\ y = 0$$
$$\mathbf{x} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}),\ y = 1$$



Linear decision boundaries

- Any other models/algorithms?

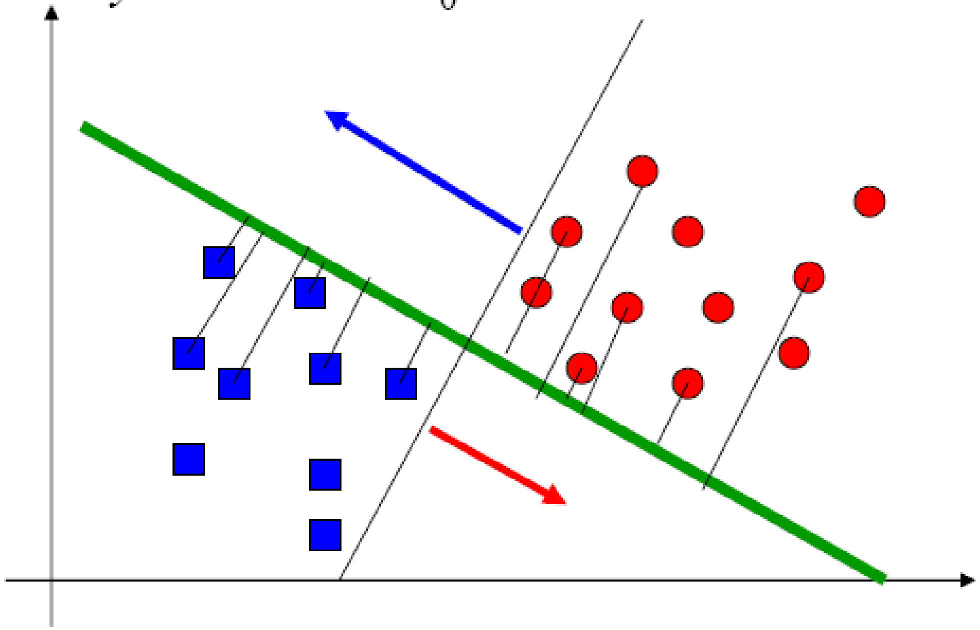


Fisher linear discriminant

- Project data into one dimension

$$y = \mathbf{w}^T \mathbf{x}$$

Decision: $y = \mathbf{w}^T \mathbf{x} + w_0 \geq 0$

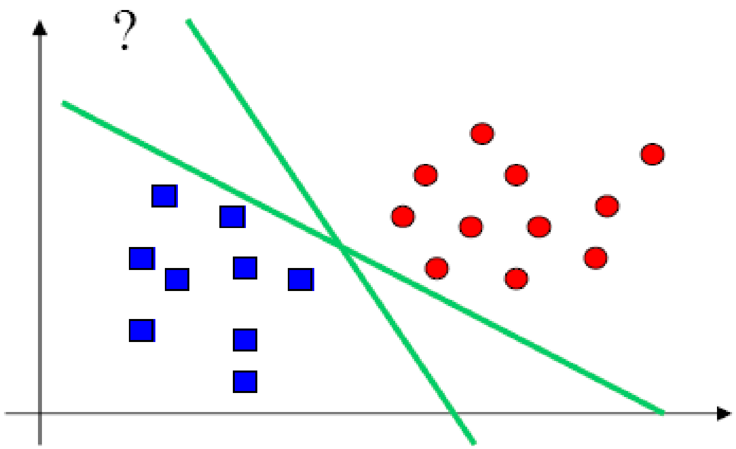


- How to find the projection line?

Fisher linear discriminant

How to find the projection line?

$$y = \mathbf{w}^T \mathbf{x}$$

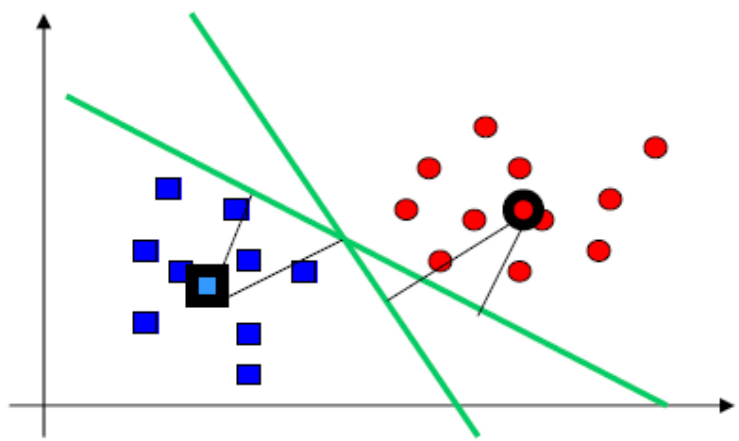


Fisher linear discriminant

Assume: $$\mathbf{m}_1 = \frac{1}{N_1}\sum_{i\in C_1}^{N_1} \mathbf{x}_i \qquad \mathbf{m}_2 = \frac{1}{N_2}\sum_{i\in C_2}^{N_2} \mathbf{x}_i$$

Maximize the difference in projected means:

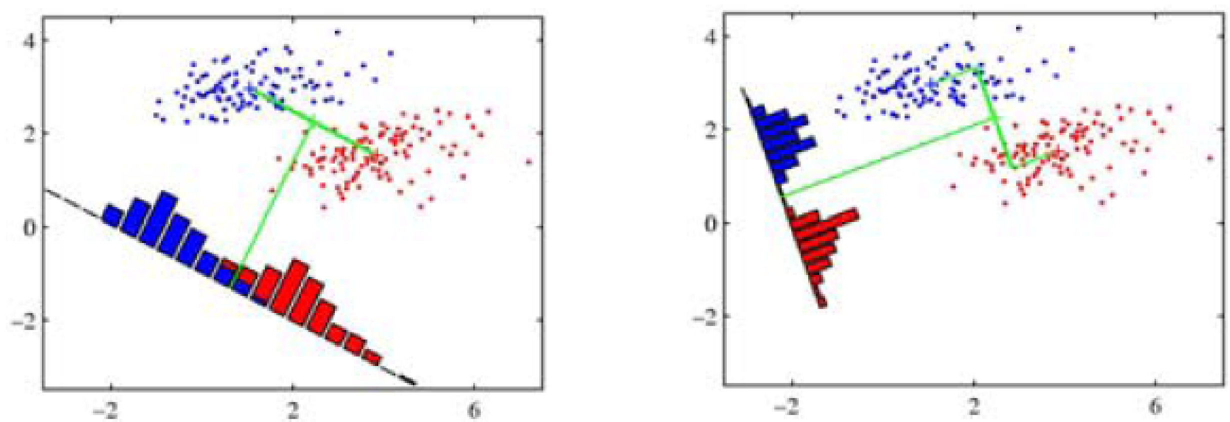
$$m_2 - m_1 = \mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)$$



Fisher linear discriminant

Problem 1: $m_2 - m_1 = \mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)$ can be maximized by increasing $\mathbf{w}$

Problem 2: variance in class distributions after projection is changed



Fisher's solution: $$J(\mathbf{w}) = \frac{m_2 - m_1}{s_1^2 + s_2^2}$$

Within class variance $$s_k^2 = \sum_{i \in C_k} (y_i - m_k)^2$$

Fisher linear discriminant

Error:

$$J(\mathbf{w}) = \frac{m_2 - m_1}{s_1^2 + s_2^2}$$

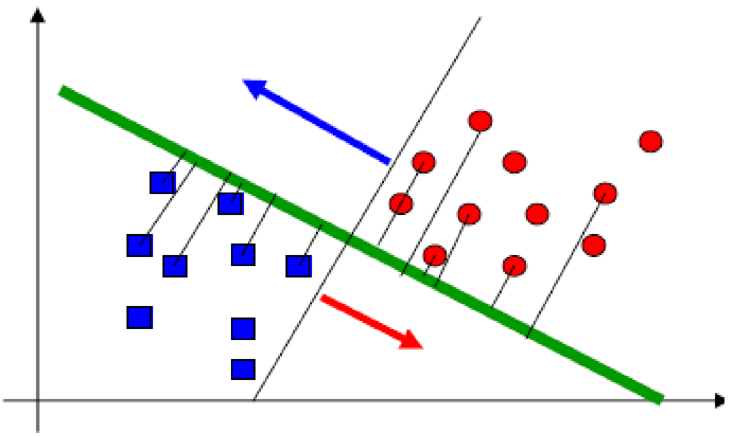
Within class variance after the projection

$$s_k^2 = \sum_{i \in C_k} (y_i - m_k)^2$$

Optimal solution:

$$\mathbf{w} \approx \mathbf{S}_\mathbf{w}^{-1}(\mathbf{m}_2 - \mathbf{m}_1)$$

$$\mathbf{S}_\mathbf{w} = \sum_{i \in C_1} (\mathbf{x}_i - \mathbf{m}_1)(\mathbf{x}_i - \mathbf{m}_1)^T + \sum_{i \in C_2} (\mathbf{x}_i - \mathbf{m}_2)(\mathbf{x}_i - \mathbf{m}_2)^T$$



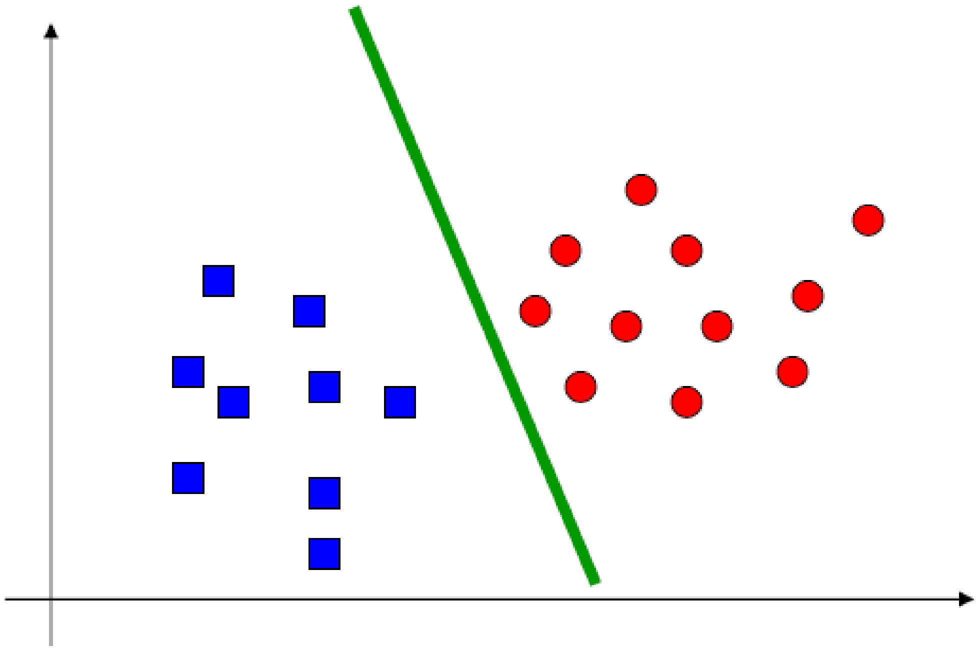
Linearly separable classes

There is a **hyperplane** that separates training instances with no error

Hyperplane:

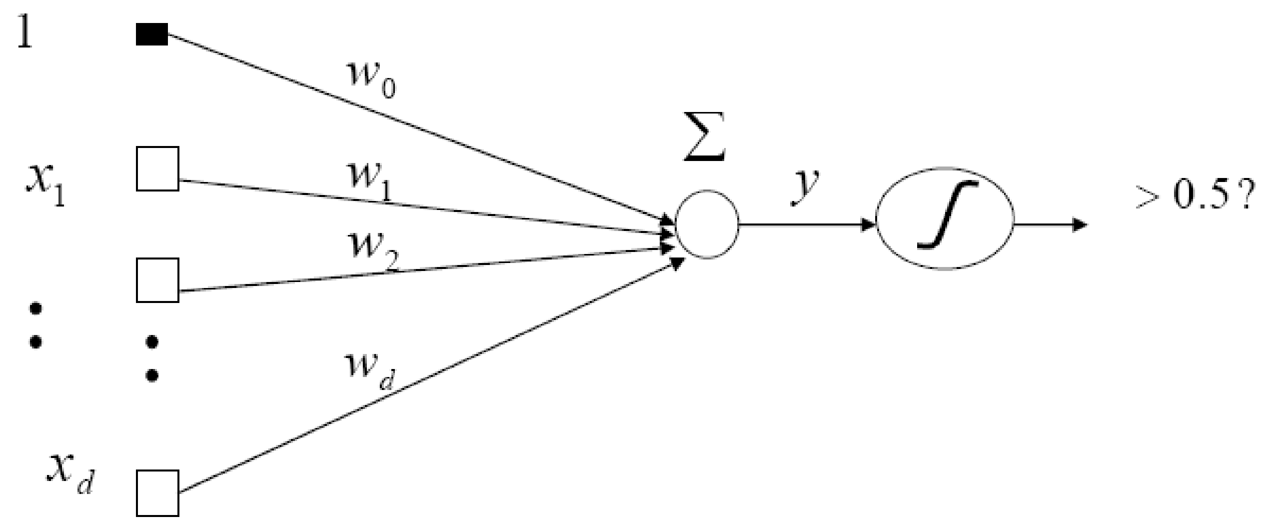
$$\mathbf{w}^T\mathbf{x} + w_0 = 0$$

Class (+1) $\mathbf{w}^T\mathbf{x} + w_0 > 0$
Class (-1) $\mathbf{w}^T\mathbf{x} + w_0 < 0$



Algorithms for linearly separable set

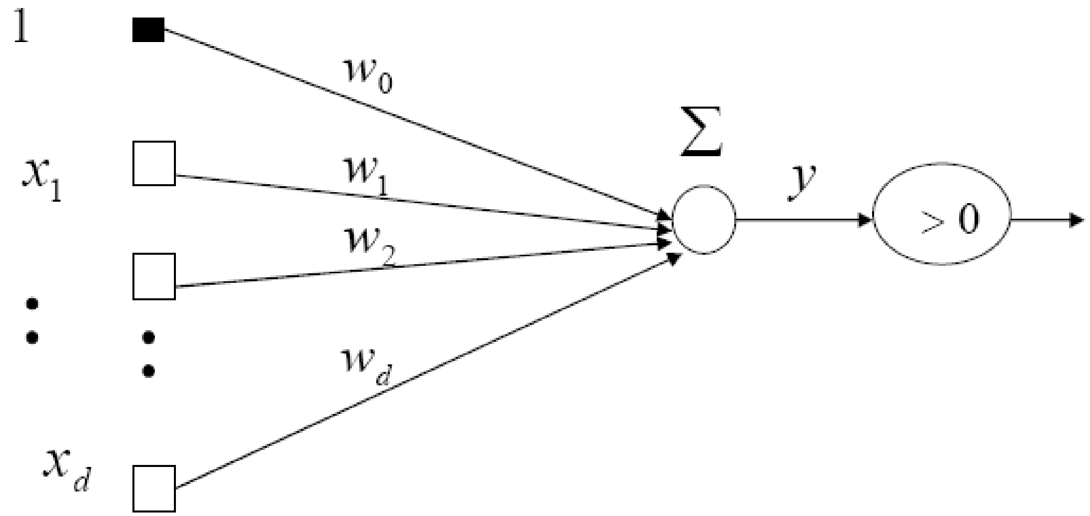
- **Separating hyperplane** $\mathbf{w}^T\mathbf{x} + w_0 = 0$



- We can use **gradient methods** or Newton Rhapson for sigmoidal switching functions and learn the weights
- Recall that we learn the linear decision boundary

Algorithms for linearly separable set

- **Separating hyperplane** $\mathbf{w}^T\mathbf{x} + w_0 = 0$



Algorithms for linearly separable sets

Perceptron algorithm:

- Simple iterative procedure for modifying the weights of the linear model
- Works for inputs $\mathbf{x}$ where each x_i is in [0,1]

Initialize weights $\mathbf{w}$

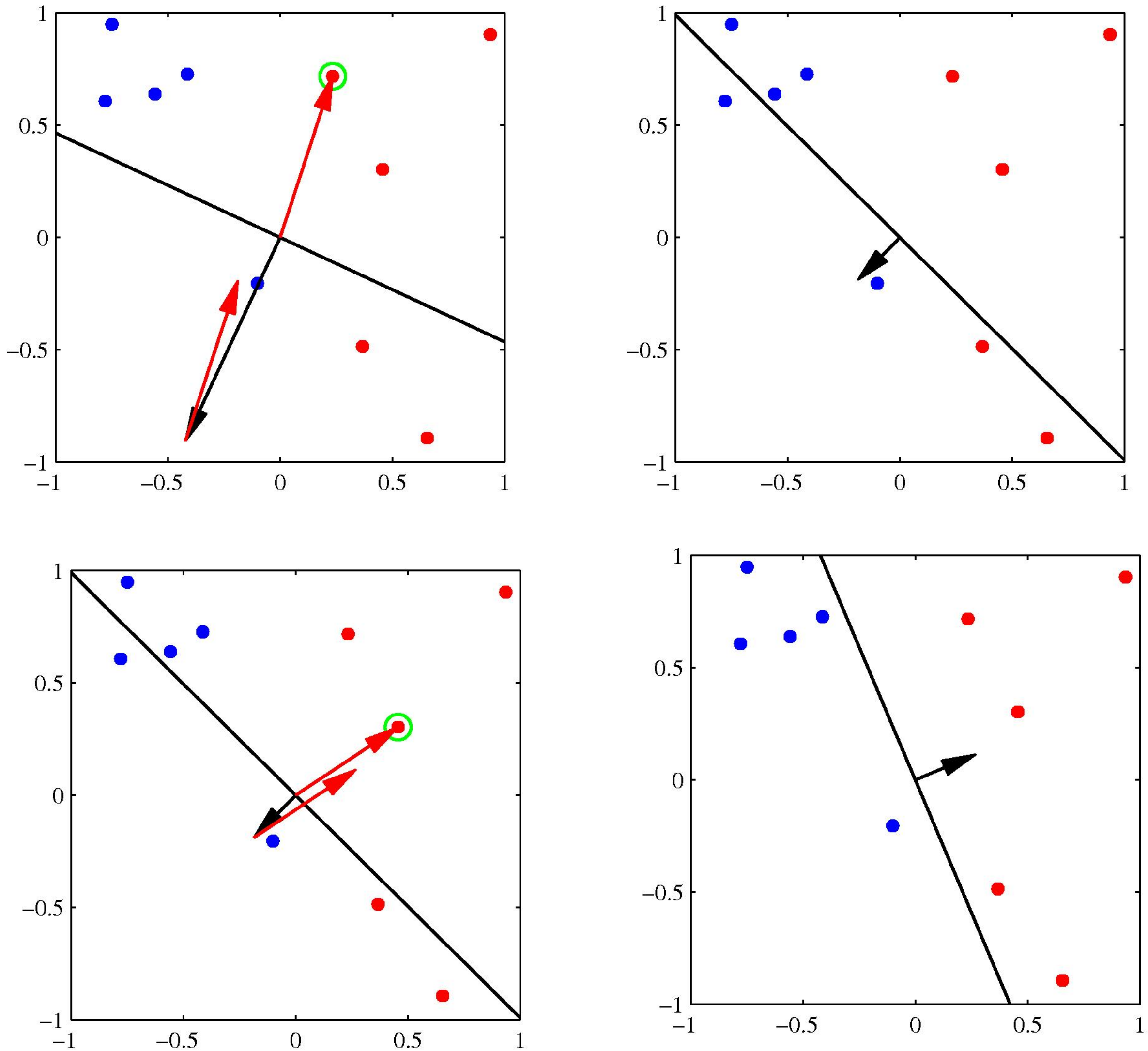
Loop through examples $(\mathbf{x}, y)$ in the dataset D

1. Compute $\hat{y} = \mathbf{w}^T \mathbf{x}$
2. If $y \neq \hat{y} = -1$ then $\mathbf{w}^T \leftarrow \mathbf{w}^T + \mathbf{x}$
3. If $y \neq \hat{y} = +1$ then $\mathbf{w}^T \leftarrow \mathbf{w}^T - \mathbf{x}$

Until all examples are classified correctly

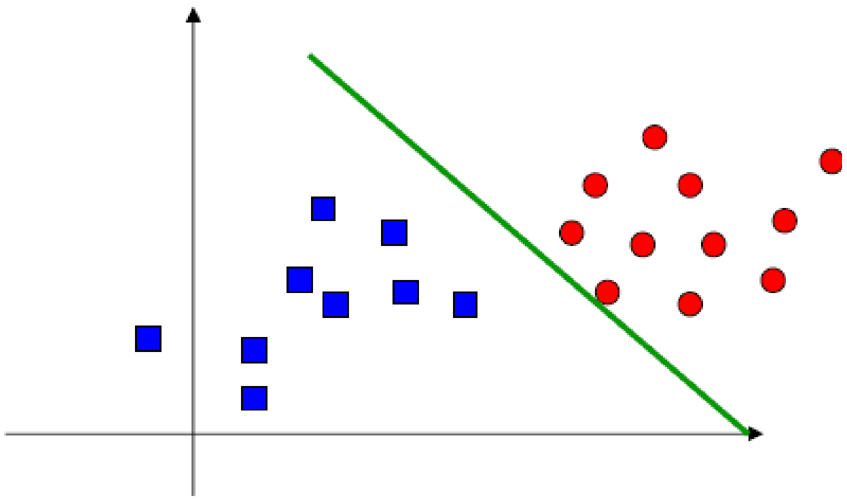
Properties:

- **guaranteed convergence if the classes are linearly separable**



−1
−0.5
0
0.5
1

Algorithms for linearly separable sets



Linear program solution:

- Finds weights that satisfy the following constraints:

$$\mathbf{w}^T \mathbf{x}_i + w_0 \geq 0 \qquad \text{For all i, such that } y_i = +1$$

$$\mathbf{w}^T \mathbf{x}_i + w_0 \leq 0 \qquad \text{For all i, such that } y_i = -1$$

Together: $y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 0$

Property: if there is a hyperplane separating the examples, the linear program finds the solution