Linear Classifiers II

Tomáš Svoboda

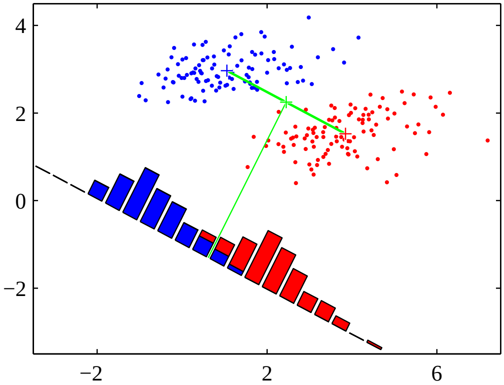
Vision for Robots and Autonomous Systems, Center for Machine Perception
Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University in Prague

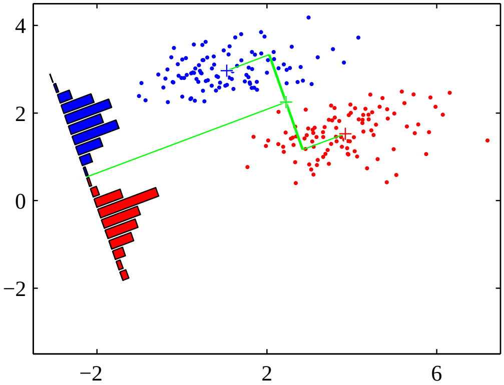
May 26, 2021

Linear Classifiers - supplement lecture

- Supplement to the lecture about learning Linear Classifiers (perceptron, ...)
- Better etalons by applying Fischer linear discriminator analysis.
- LSQ formulation of the learning task.

Fischer linear discriminant





- Dimensionality reduction
- Maximize distance between means, ...
- ...and minimize within class variance. (minimize overlap)

Figures from [1]

Projections to lower dimensions $y = \mathbf{w}^\top \mathbf{x}$

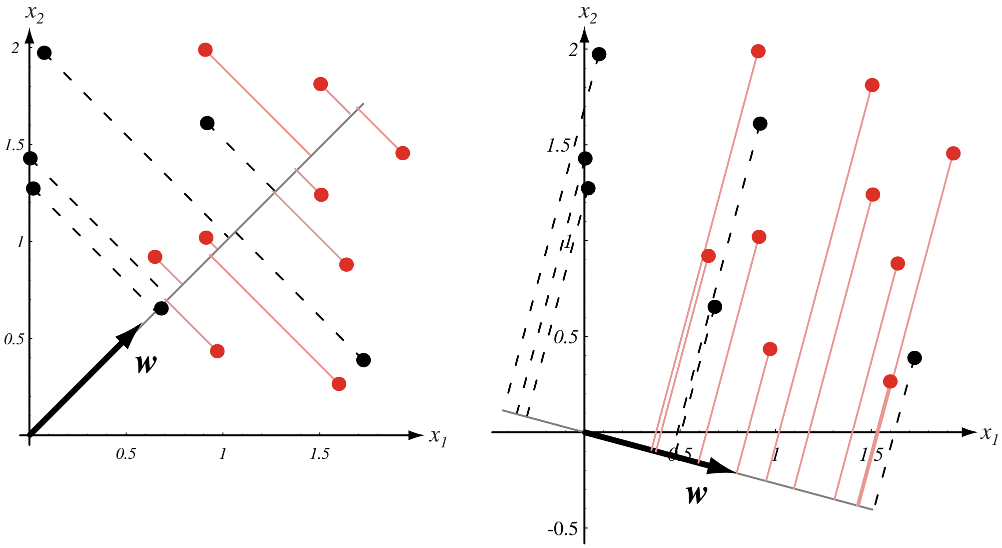


Figure from [2]

Projection to lower dimension $\mathbf{y} = \mathtt{W}^\top \mathbf{x}$

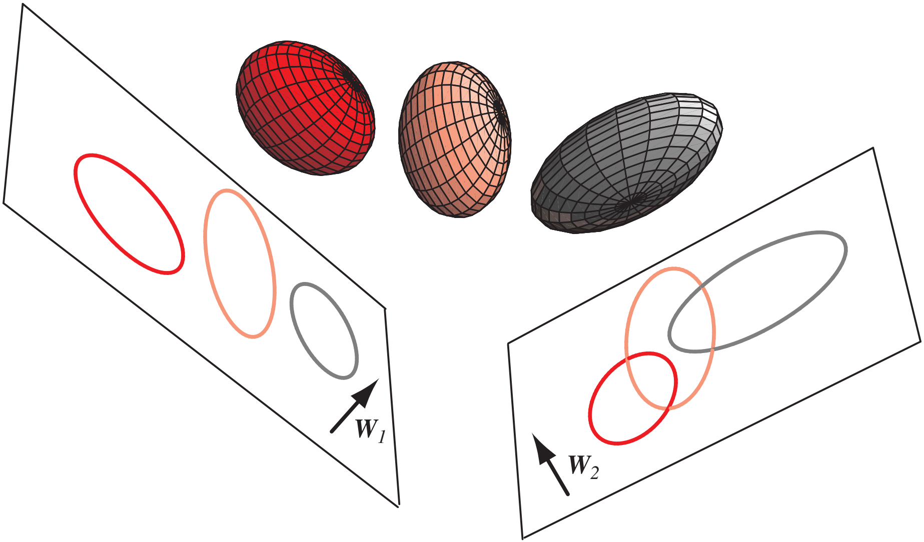
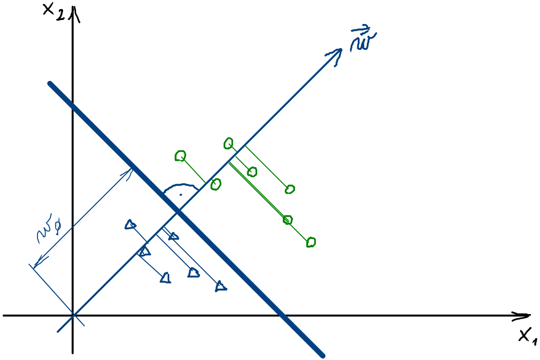
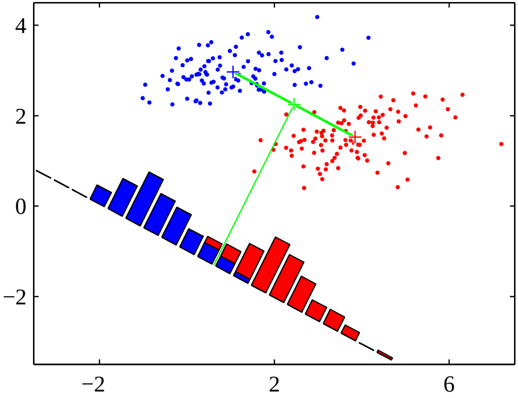


Figure from [2]

Finding the best projection $y = \mathbf{w}^\top \mathbf{x}$, $y \geq -w_0 \Rightarrow C_1$, otherwise C_2





Finding the best projection $y = \mathbf{w}^\top \mathbf{x}$, $y \geq -w_0 \Rightarrow C_1$, otherwise C_2

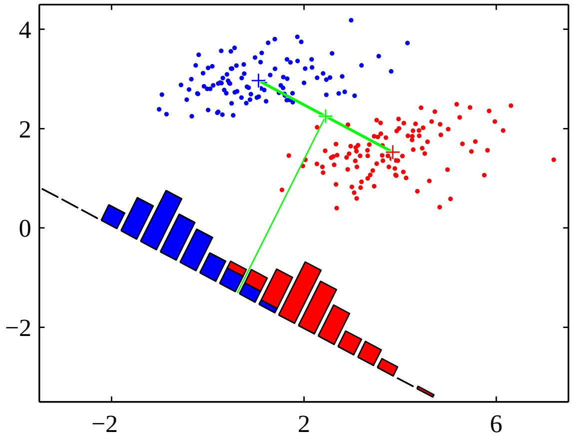
$$m_2 - m_1 = \mathbf{w}^\top(\mathbf{m}_2 - \mathbf{m}_1)$$

Within class scatter of projected samples

$$s_i^2 = \sum_{y \in C_i} (y - m_i)^2$$

Fischer criterion:

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



Finding the best projection $y = \mathbf{w}^\top \mathbf{x}$, $y \geq -w_0 \Rightarrow C_1$, otherwise C_2

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

$$s_i^2 = \sum_{y \in C_i} (y - m_i)^2$$

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

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = 0$$

$$\mathrm{S}_i = \sum_{x \in C_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^\top$$

$$\mathrm{S}_W = \mathrm{S}_1 + \mathrm{S}_2$$

$$\mathrm{S}_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^\top$$

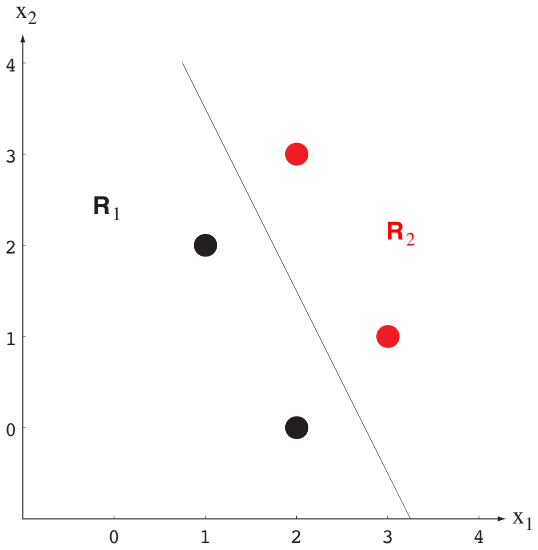
$$J(\mathbf{w}) = \frac{\mathbf{w}^\top \mathrm{S}_B \mathbf{w}}{\mathbf{w}^\top \mathrm{S}_W \mathbf{w}}$$

LSQ approach to linear classification

$$\mathbf{w} = \begin{bmatrix} w_0 \\ \mathbf{w} \end{bmatrix}$$

$$\mathtt{X}\mathbf{w} = \mathbf{b}$$

$$J(\mathbf{w}) = \|\mathtt{X}\mathbf{w} - \mathbf{b}\|^2$$



LSQ approach, better margins b?

$$\mathtt{X} = \begin{bmatrix} \mathtt{1}_1 & \mathtt{X}_1 \\ -\mathtt{1}_2 & -\mathtt{X}_2 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} \frac{n}{n_1}\mathtt{1}_1 \\ \frac{n}{n_2}\mathtt{1}_2 \end{bmatrix}$$

References I

Further reading: Chapter 4 of [1], or chapter 3 and 5 of [2].

[1] Christopher M. Bishop.
Pattern Recognition and Machine Learning.
Springer Science+Bussiness Media, New York, NY, 2006.
PDF freely downloadable.

[2] Richard O. Duda, Peter E. Hart, and David G. Stork.
Pattern Classification.
John Wiley & Sons, 2nd edition, 2001.