Linear Models for Classification

Teemu Mutanen 5.2.2007

Classification problem

Why?

Predictive power

Aid to communication

Failures might be interesting

Simple - Easy

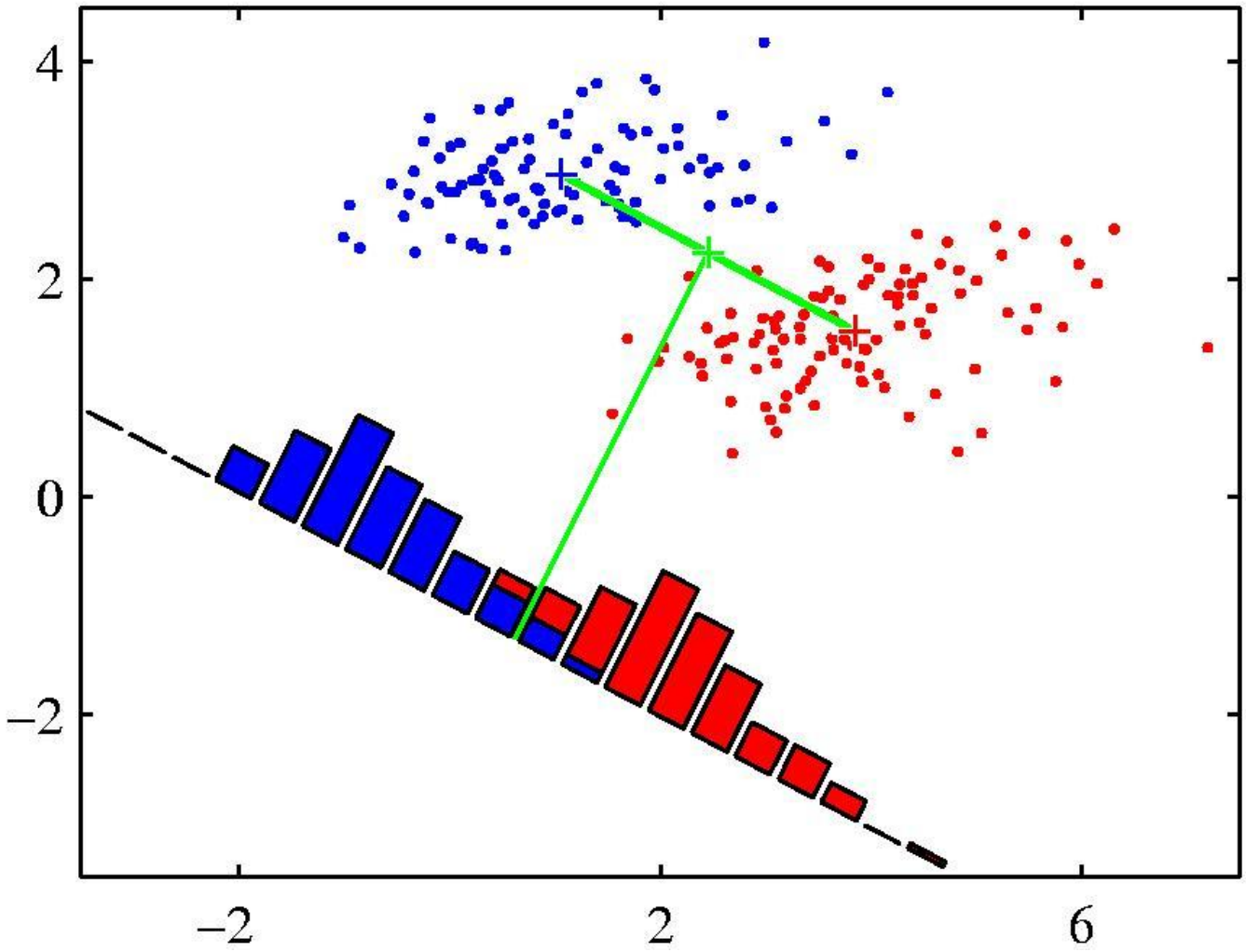
Linear discriminant function

Decision surfaces are hyperplanes

input: vector x
output: assignments

Disjoint classes – one class and only one class

Fisher ’s linear discriminant

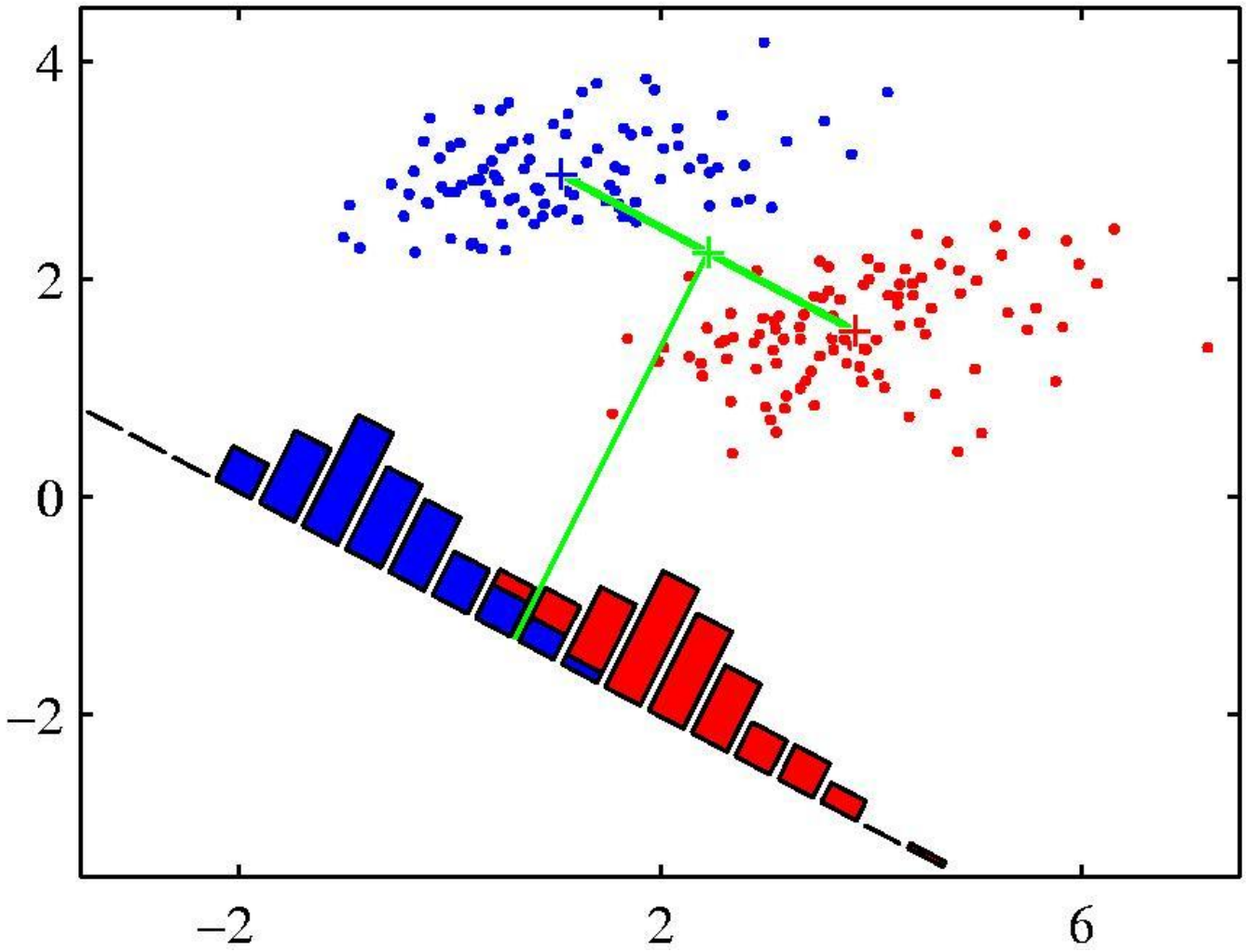


Fisher's linear discriminant

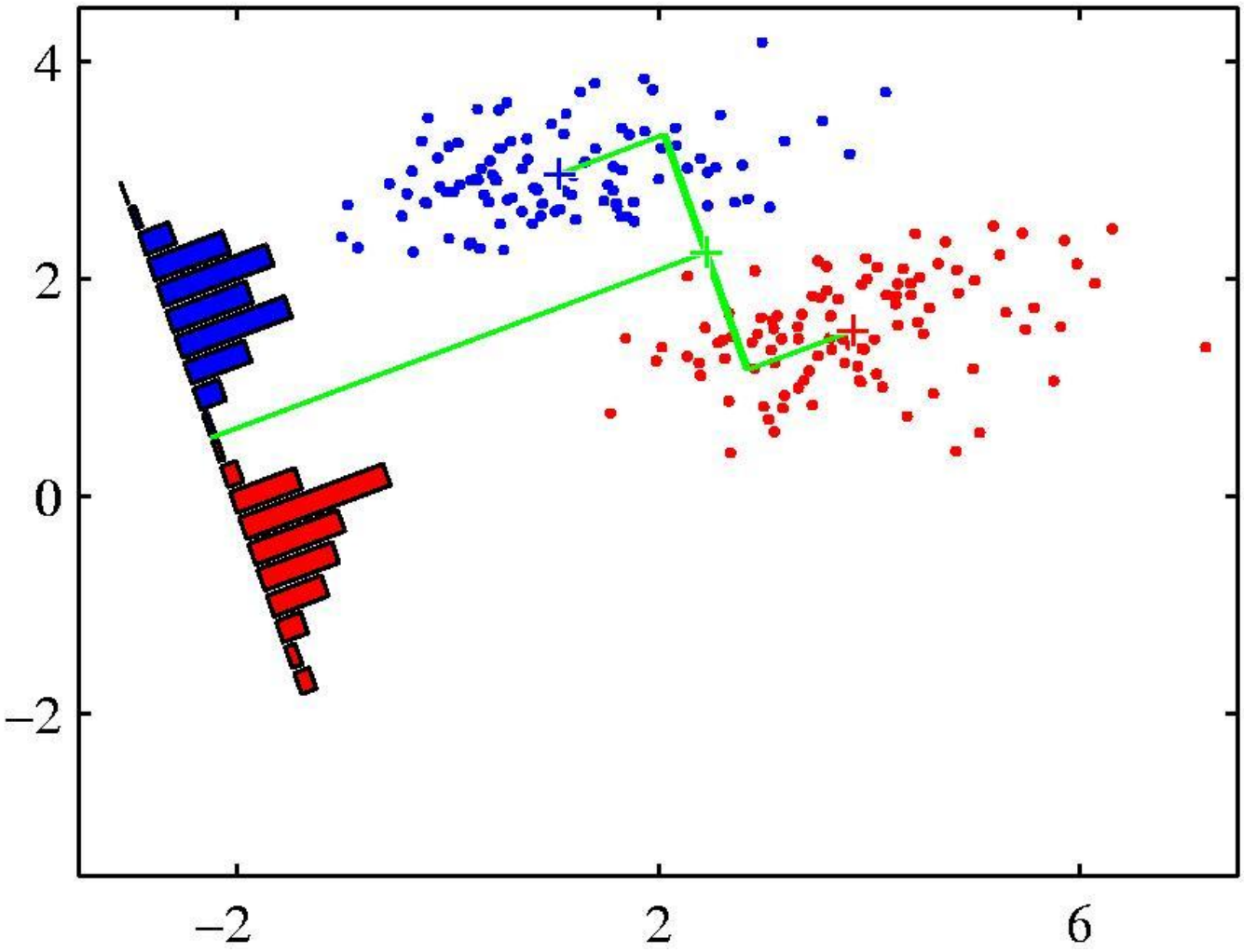
Large separation between the projected class means while also giving a small variance within each class.

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

Fisher ’s linear discriminant



Fisher ’s linear discriminant



The Perceptron Algorithm

Frank Rosenblatt (1962)

Two class model

Input vector x is transformed to give a feature vector $\phi(x)$

$$y(x) = f(w^T \phi(x))$$

The Perceptron Algorithm

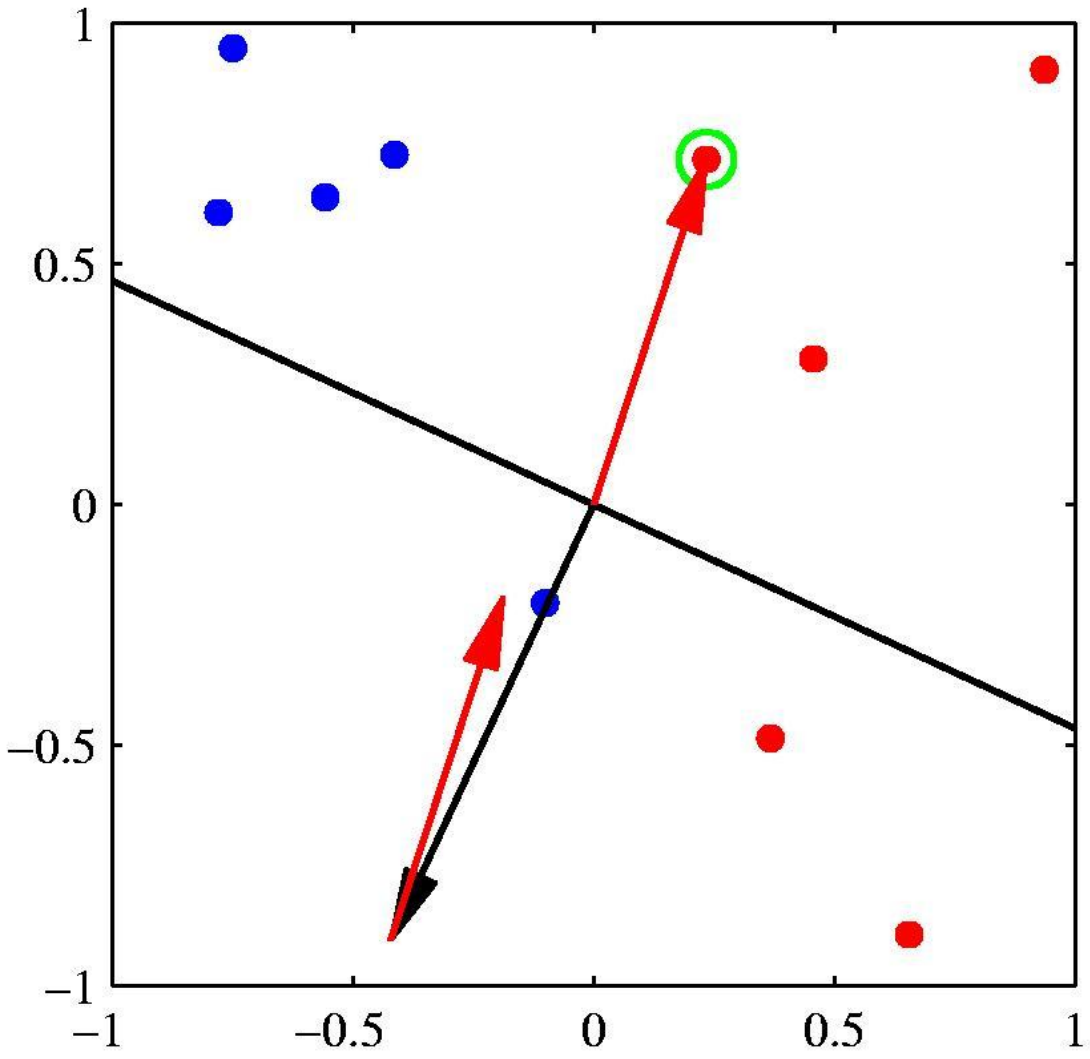
Missclassified patterns

$$0 \quad \text{or} \quad -w^T \phi(x_n) t_n$$

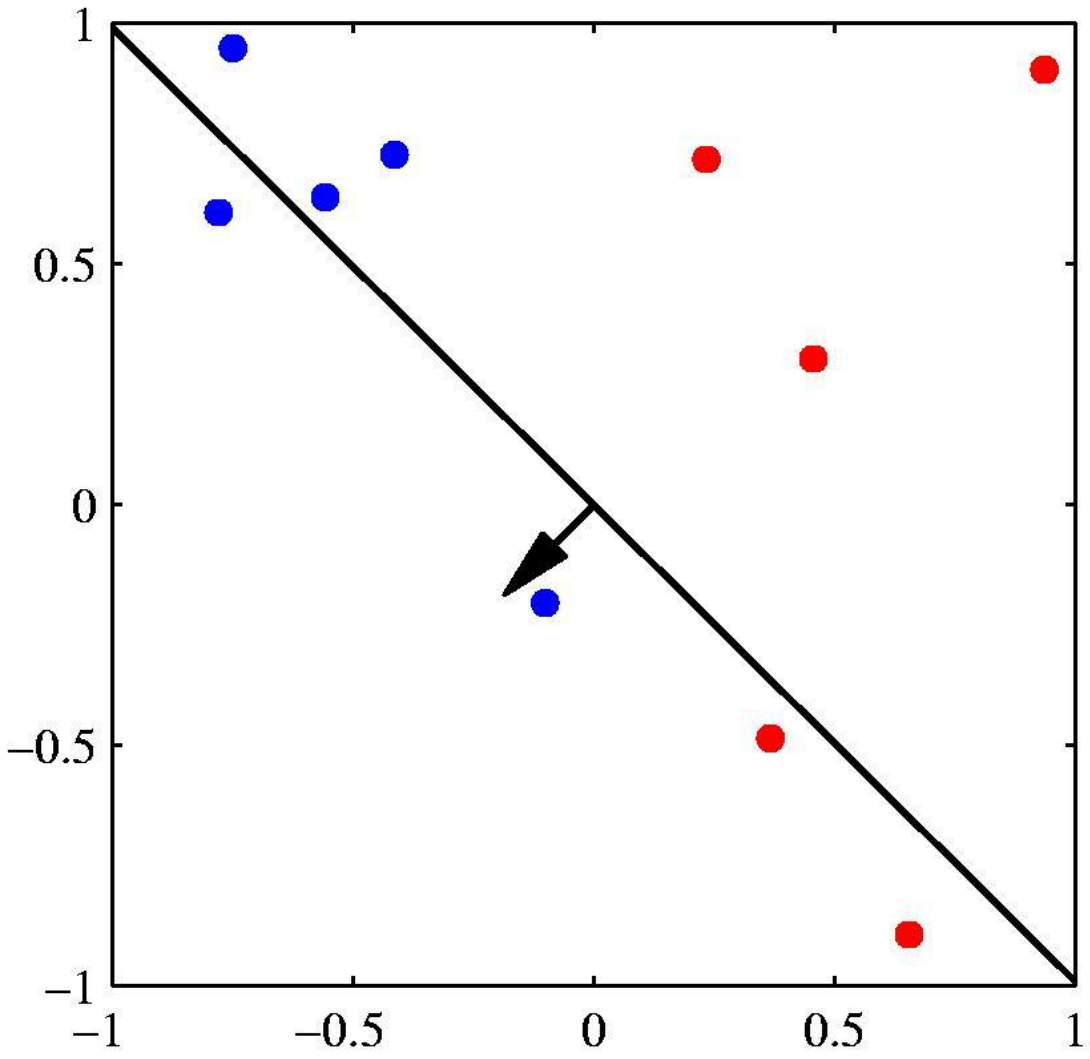
The perceptron criterion:

$$E_p(w) = -\sum_{n \in \mathrm{M}} w^T \phi_n t_n$$

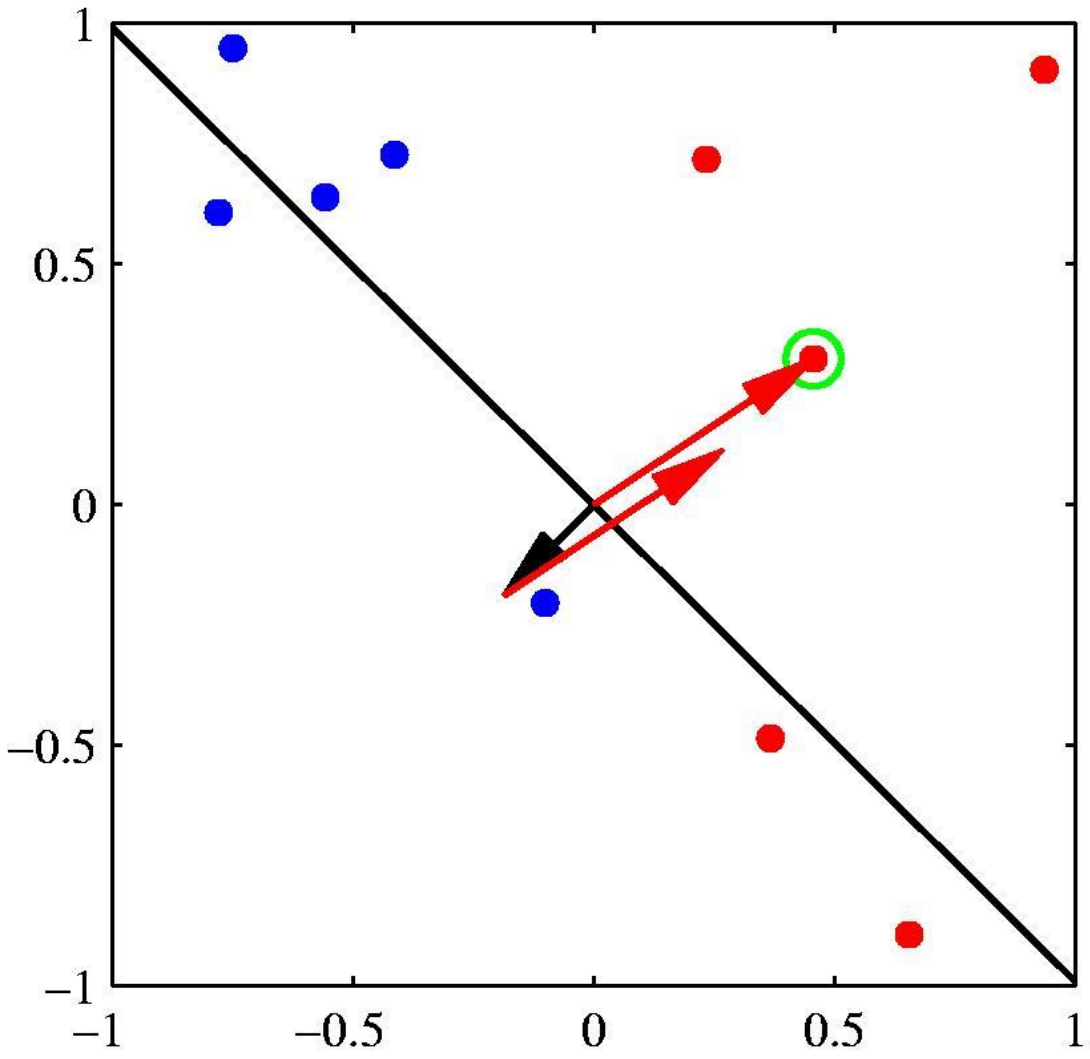
The Perceptron Algorithm



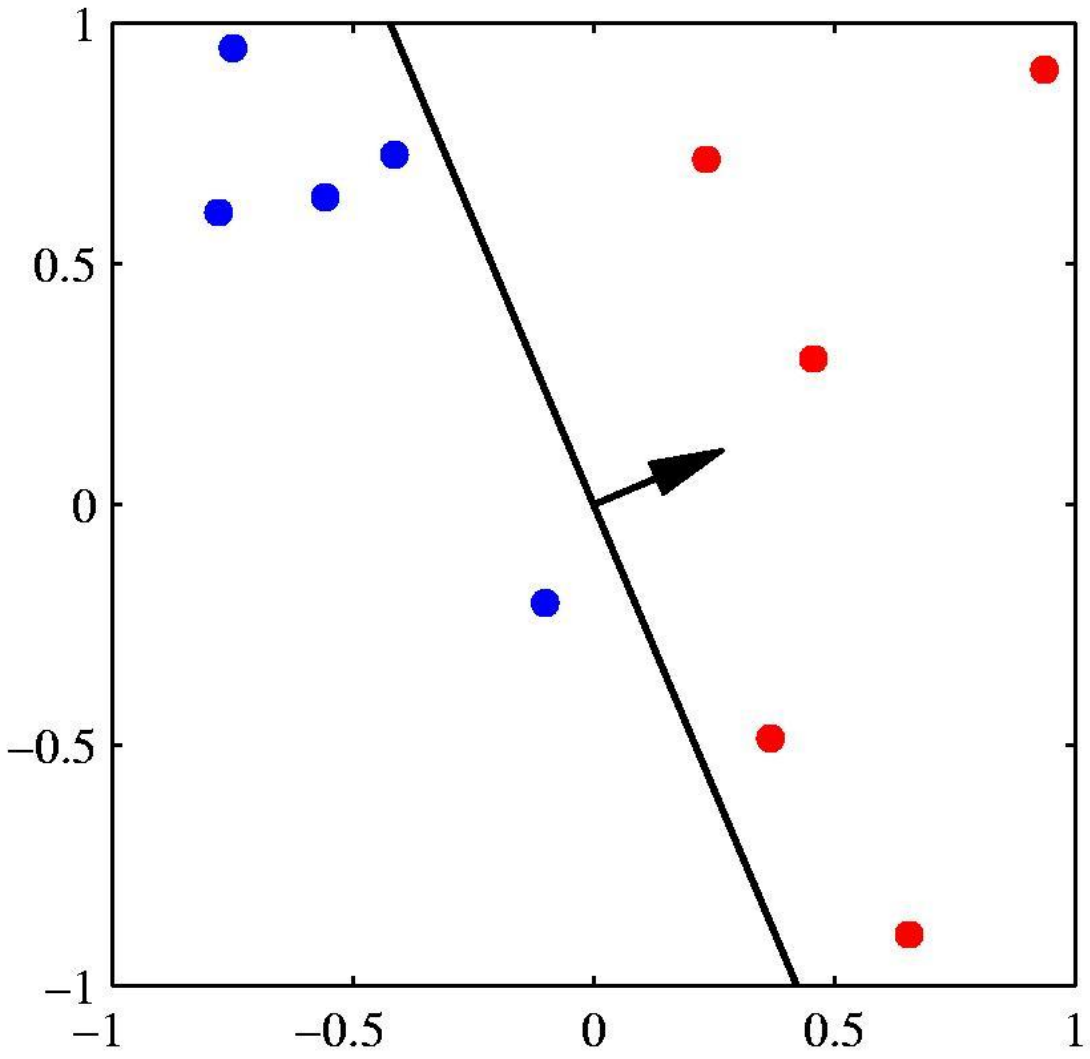
The Perceptron Algorithm



The Perceptron Algorithm



The Perceptron Algorithm



Probabilistic Generative Models

Class-conditional densities $p(x|C_k)$

Class priors $p(C_k)$

Posterior probabilities $p(C_k|x)$

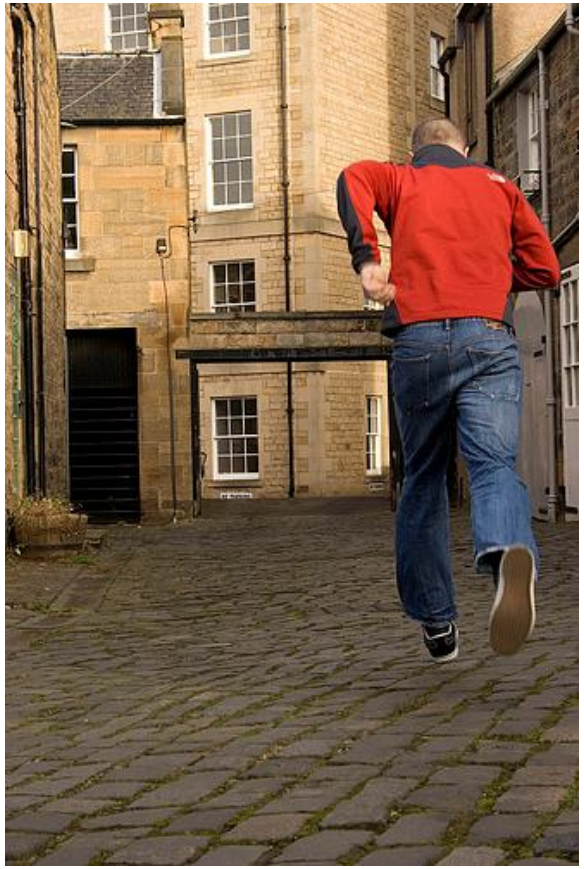
Two class case:

$$p(C_1|x) = \frac{p(x|C_1)p(C_1)}{p(x|C_1)p(C_1) + p(x|C_2)p(C_2)}$$

Take the money and run!

Consider the following exercises:

4.9*, 4.11**, and 4.16*



Thank you for your attention!

photos © winding road, greentheory, dirtyone30, and colinj @ flickr