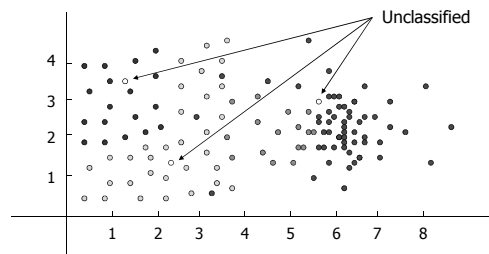


## k Nearest Neighbor (kNN) Classification Example



◆ What is the class of each unclassified sample??

## kNN Classification

- ◆ Find the  $k$  nearest neighbors of the test sample
- ◆ Classify the test sample with the majority class of its  $k$  nearest neighbors

## About kNN

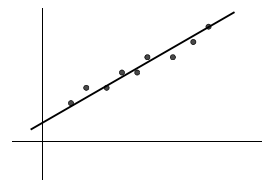
- ◆ Similarity/distance measures
  - More on this when we talk about clustering
- ◆ Index structures, e.g. k-d tree
- ◆ Local decision – susceptible to noise
- ◆ Error rate  $\leq (2 * \text{Bayes Error Rate})$  if  $k=1$  and  $n \rightarrow \infty$

## Predicating Continuous Values

- ◆ Regression methods
  - Linear regression
  - Non-linear regression
- ◆ Other methods
  - Some classification methods can be used or adapted to predict continuous values, e.g. ANN and kNN

## Linear Regression

- ◆ Record  $(x, y)$ 
  - $x$ : predictor variable
  - $y$ : response variable
- ◆ Model
  - $y = w_0 + w_1x$



## Linear Regression Using Least-Squares Method

$$w_1 = \frac{\sum_{i=1}^{|D|} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{|D|} (x_i - \bar{x})^2}$$

$$w_0 = \bar{y} - w_1 \bar{x}$$

## Multiple Linear Regression

◆ Record  $(x_1, \dots, x_n, y)$

◆ Model:

$$y = w_0 + \sum_{i=1}^n w_i x_i$$

## Readings

◆ Textbook 6.9 and 6.11