Supervised Learning

Prof. Dr. Ronny Hartanto

HOCHSCHULE RHEIN-WAAL
Rhine-Waal University of Applied Sciences

Supervised Learning

Objective:

Given some input data $x \epsilon F$, predict the target values $y \epsilon L$ according to some performance measure P.

The input data space F is also called "feature space".

Supervised Learning has two phases:

1. Training period:
   Receive a set of input data with corresponding target values:
   $$(x_i, y_i), i = 1, \dots, n$$
2. Test period:
   Receive a set of input data **without** target values.
   $$x_i , i = 1, \dots, m$$

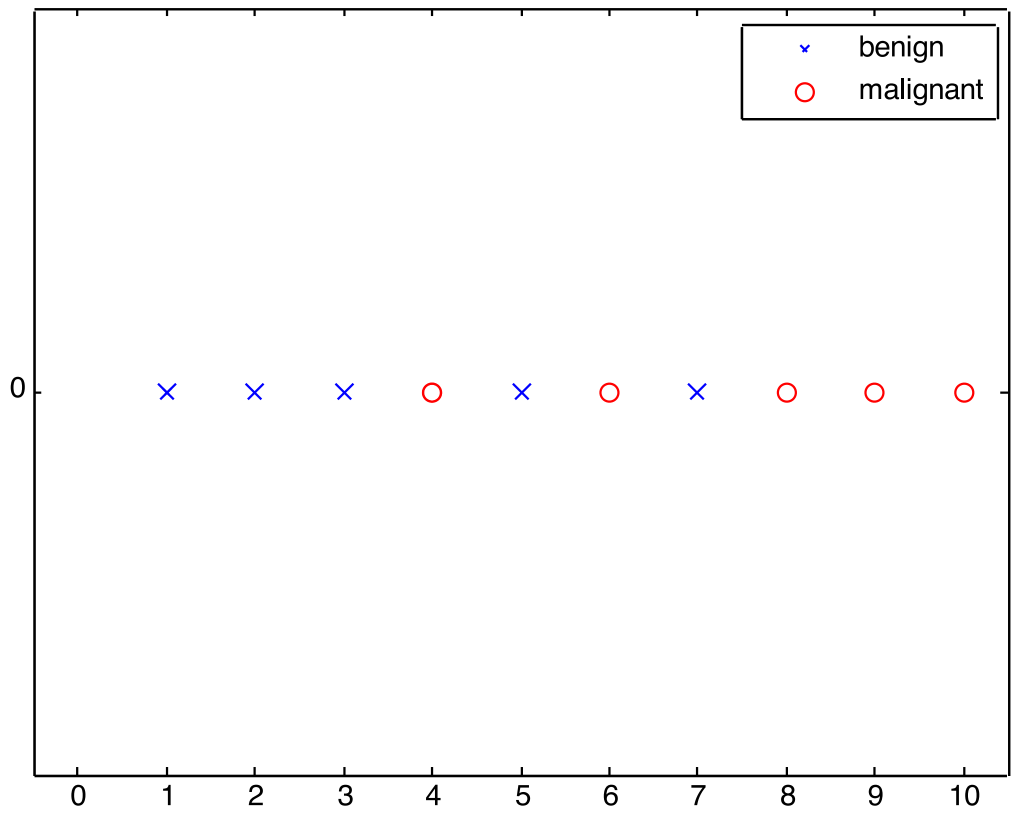
Supervised Learning

Example:

Wisconsin breast cancer database (link)

F: Tumor size (integer numbers from 1 to 10)

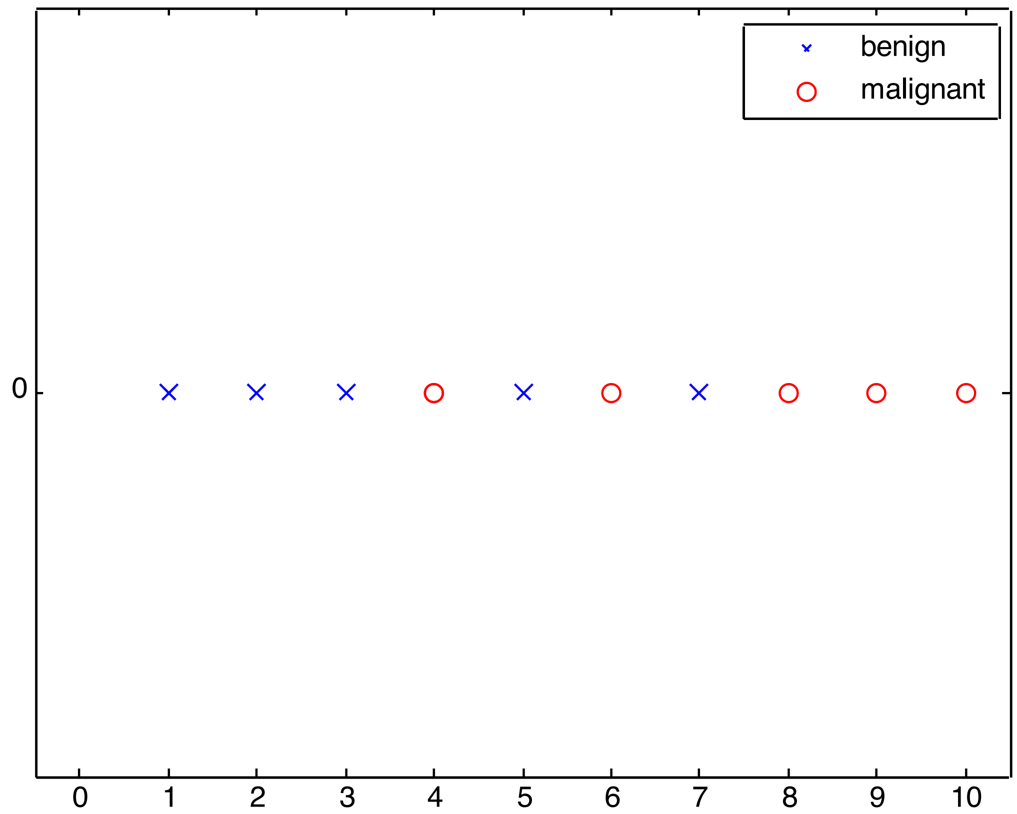
L: {"benign", "malignant"} or { x, o}, or {0,1}.



Supervised Learning

Exercise:

Write down the 10 samples that you see here into a table (Example: the leftmost data point is (1,0)).



HOCHSCHULE RHEIN-WAAL
Rhine-Waal University of Applied Sciences

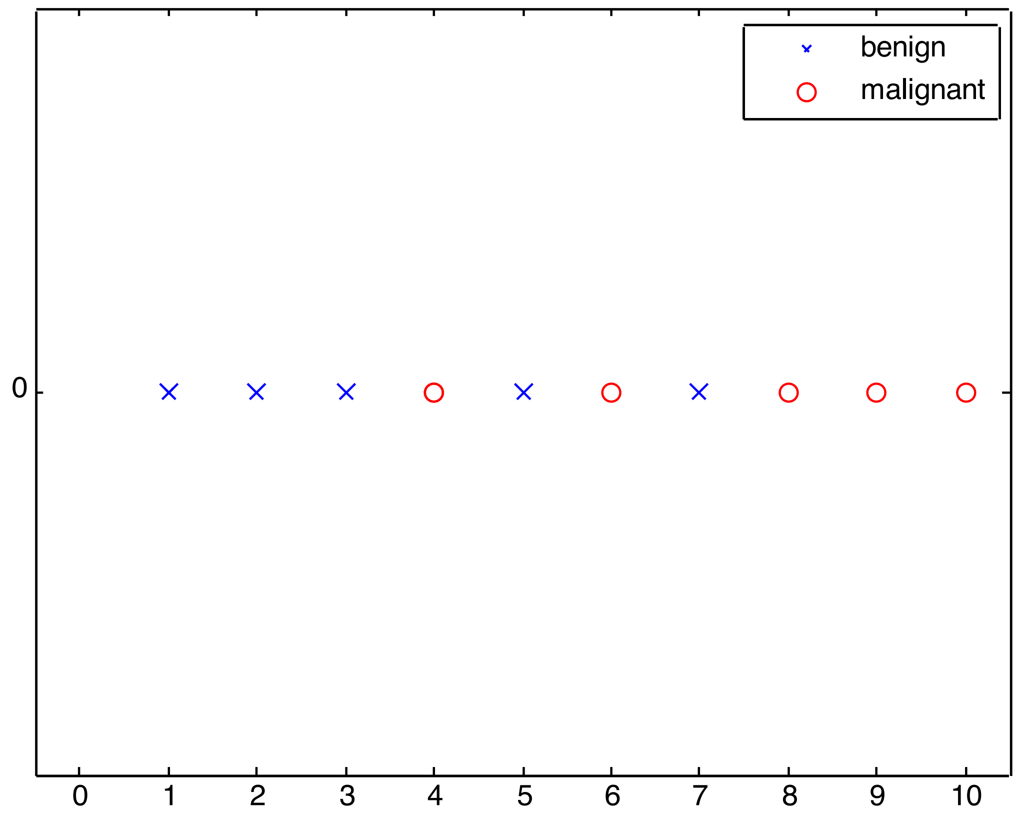
Supervised Learning

Classification:

If the target space consists of finite discrete values, the learning task is also called classification:

"Assign the most probable class label"

A function $f : F \rightarrow L$ is called a *classifier function* in this case.



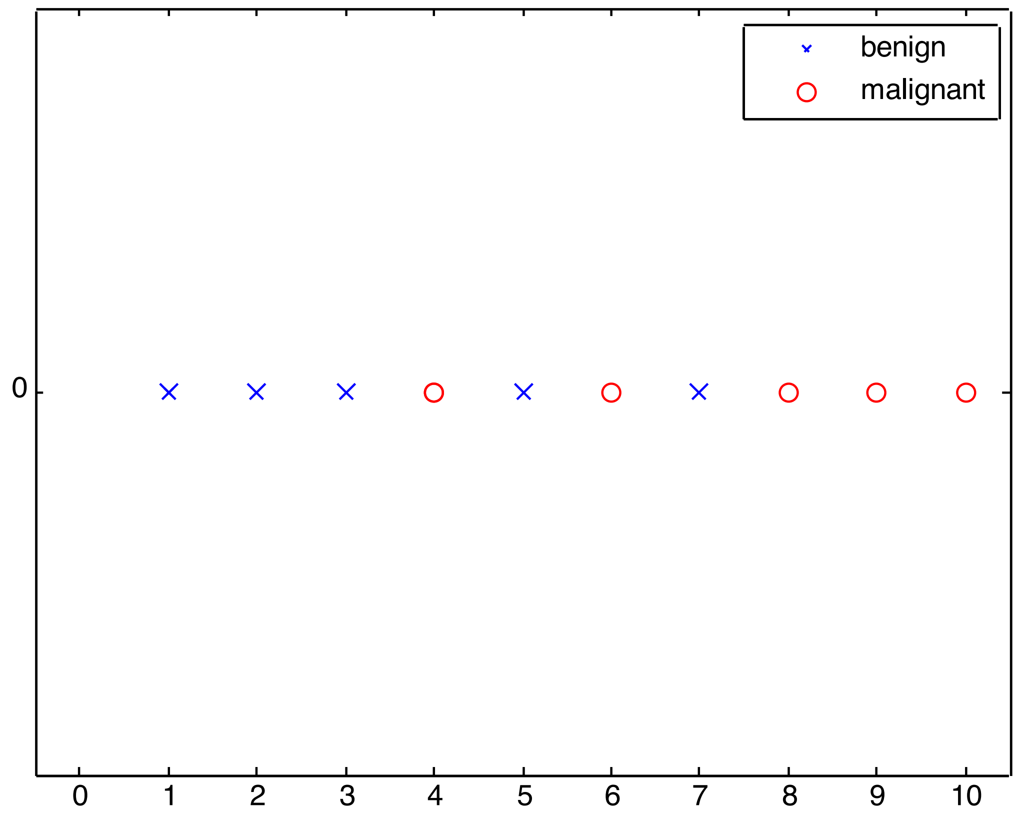
Supervised Learning

Test period:

Suppose you have a new data point (e.g. the size of an unknown new tumor is 7).

What y value do you predict?

Why?

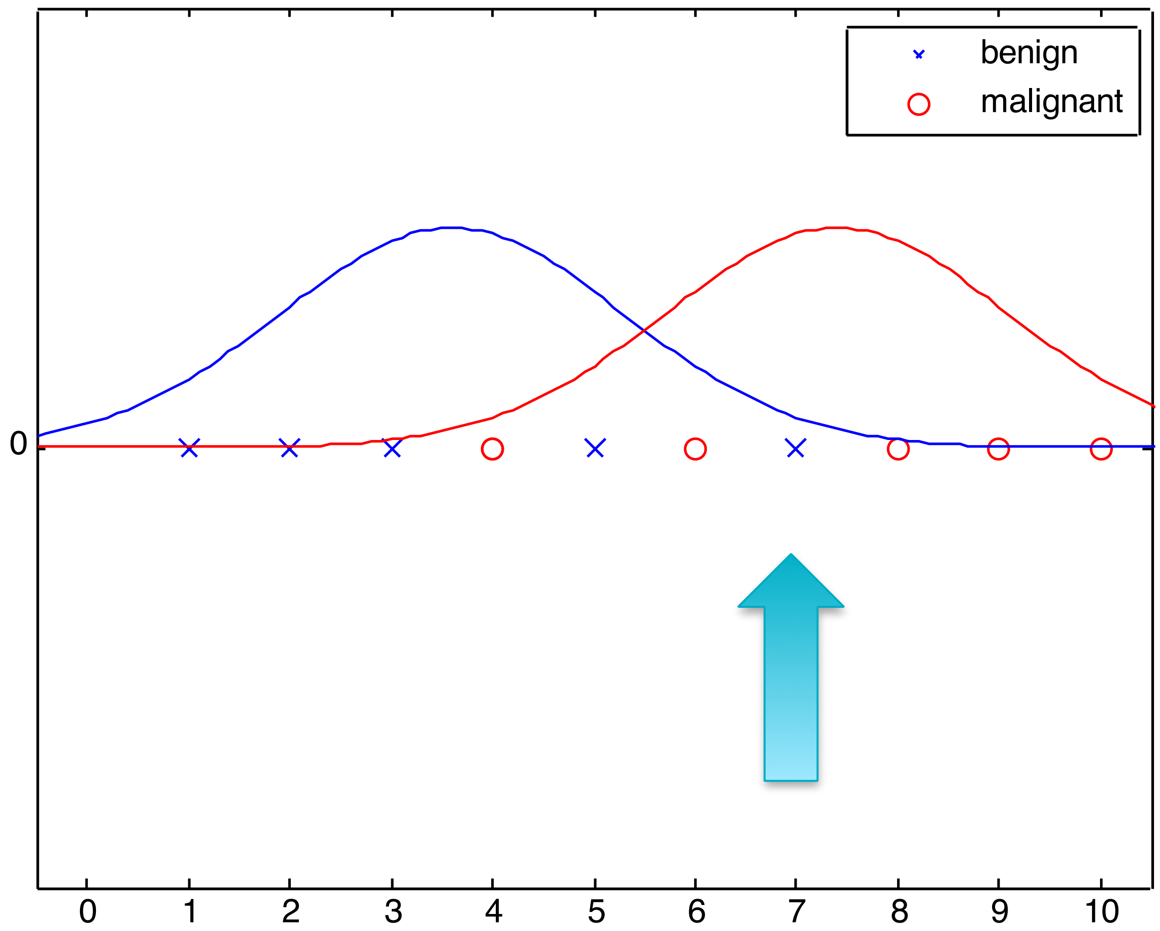


Supervised Learning

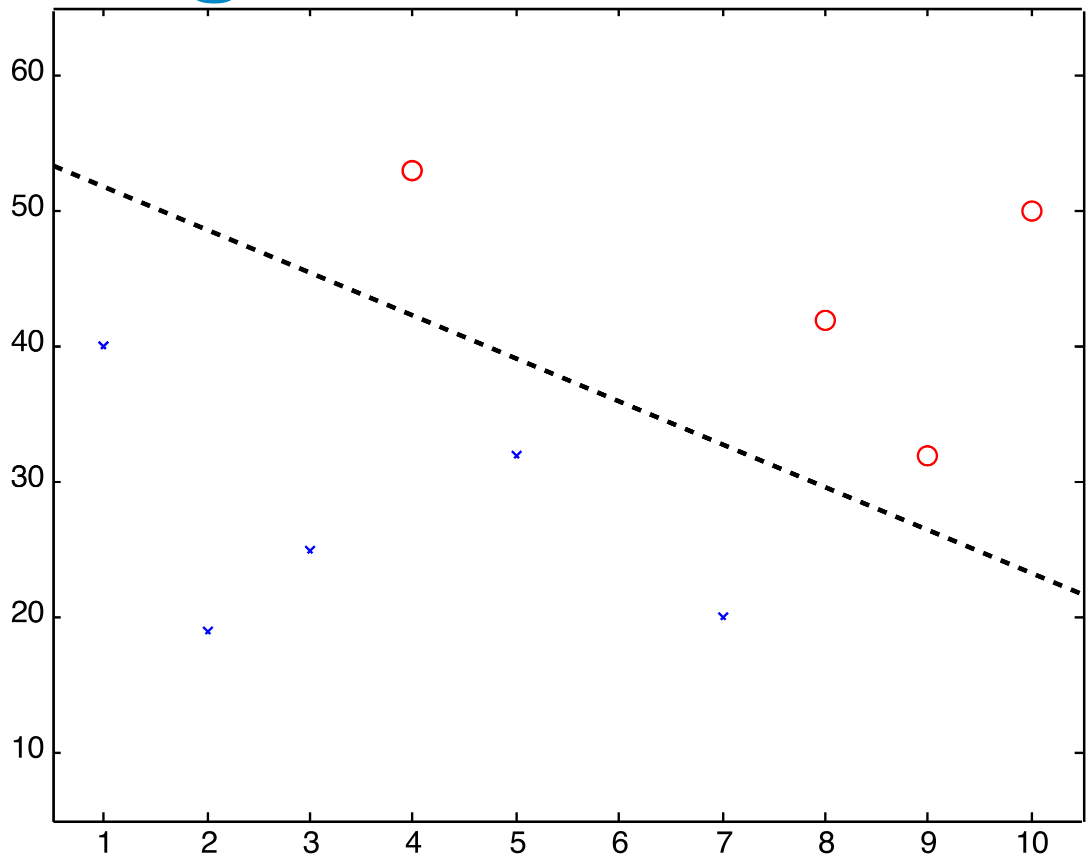
We can use findings and models from statistics.

Assumption: tumor sizes of each class are normally distributed, with equal variance.

According to these probability distributions, we should opt for "malignant".



Supervised Learning



Also in multiple dimensions:
The assumptions
1. Normal distributions
2. Equal Covariance matrices
lead to a ***linear*** classifier function.

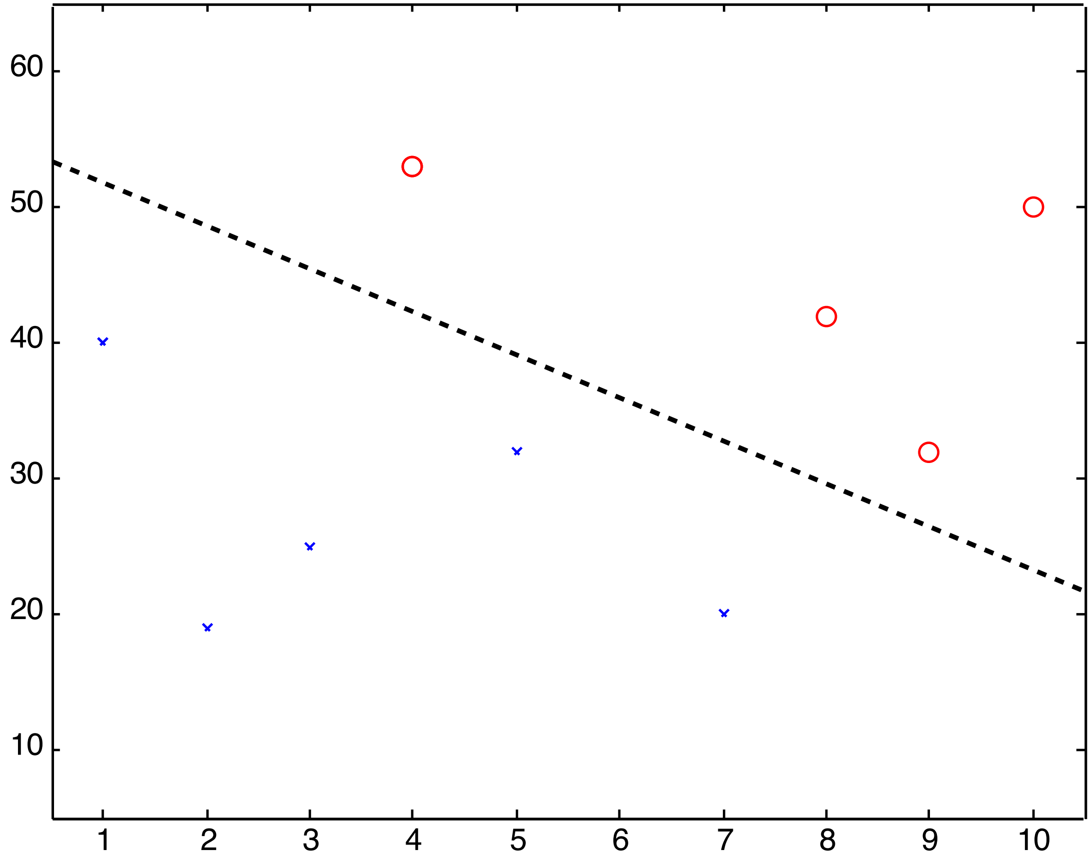
$$f(x) = \begin{cases} 1, if\ a_1 x_1 + \cdots + a_m x_m + c > 0 \\ 0, \qquad otherwise. \end{cases}$$

Supervised Learning

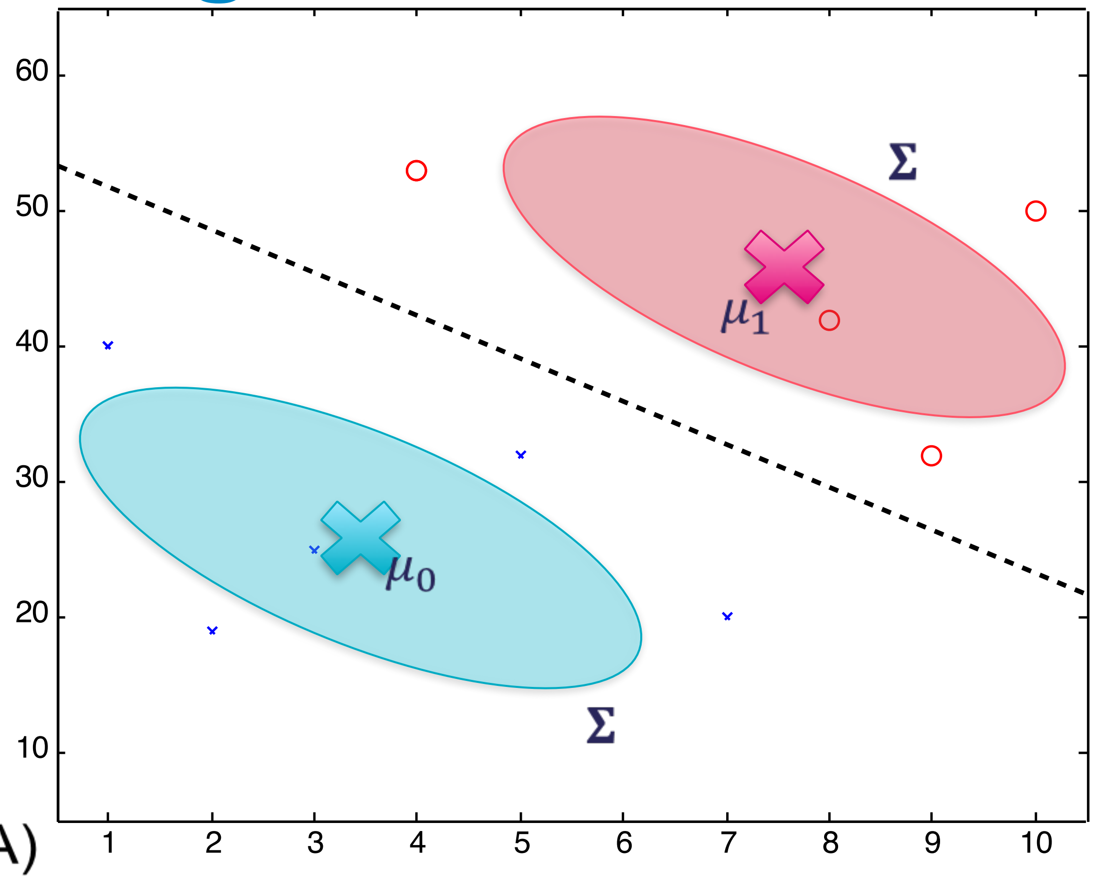
Observation:

Using more dimensions may lower the classification error.

The [Wisconsin breast cancer dataset]() contains multiple features, such as clump thickness, uniformity of cell size, cell shape,…



Supervised Learning

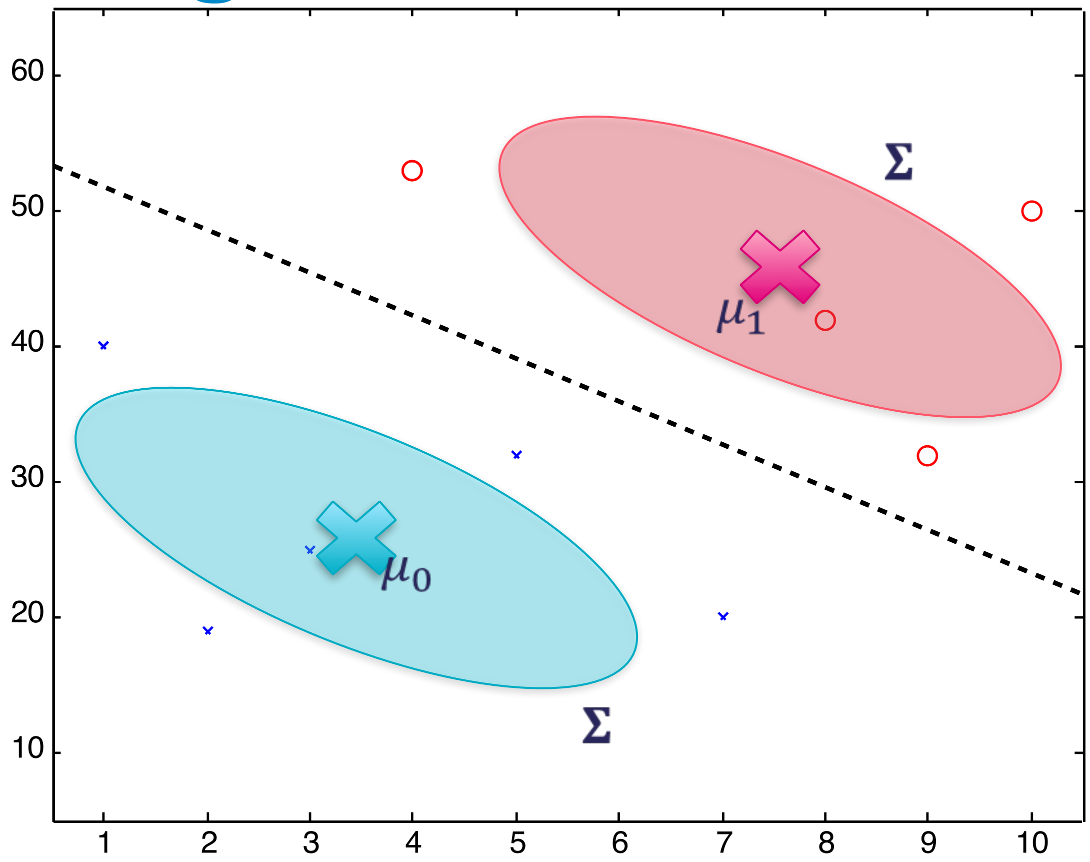


Linear Discriminant Analysis (LDA) Algorithm:

- Estimation of the class means μ_0 and μ_1
- Estimation of the "within-class" scatter matrix $\Sigma = \frac{\Sigma_0+\Sigma_1}{2}$, where $\Sigma_j = \frac{1}{|D_j-1|}\sum_{i \epsilon D_j}(x_i - \mu_j)(x_i - \mu_j)^t$

HOCHSCHULE RHEIN-WAAL
Rhine-Waal University of Applied Sciences

Supervised Learning



Linear Discriminant Analysis (LDA)

Algorithm:

- The separating hyperplane has the equation

$$w^t(x - x_0) = 0$$

where

$$w = \Sigma^{-1}(\mu_1 - \mu_0)$$

$$x_0 = \frac{(\mu_1 + \mu_0)}{2}$$

Supervised Learning

Exercise:

Suppose the training data on the right are given.

1. Draw the data into a diagram.
2. Calculate the class means μ_0 and μ_1
3. Derive the classifier parameters w and x_0
4. Draw the separating hyperplane.
5. How would you label the point (3,2) ?

x_1	x_2	y
1	*1.5*	*0*
2	*4*	*0*
6	*4*	*0*
2	*0*	*+1*
4.5	*0.5*	*+1*
6	*0.5*	*+1*
6	*2*	*+1*

Hint: The inverse of the scatter matrix is

$$\Sigma^{-1} = \begin{pmatrix} 0.34 & -0.43 \\ -0.43 & 1.26 \end{pmatrix}$$