Multi-Modal Human-Computer Interaction

Attila Fazekas

Attila.Fazekas@inf.unideb.hu

IPGD
Image Processing Group of Debrecen

Roadmap

- Basic model

Roadmap

- Basic model
- Related tasks

Roadmap

- Basic model
- Related tasks
- Face detection, facial gestures recognition

Roadmap

- Basic model
- Related tasks
- Face detection, facial gestures recognition
- Techniques

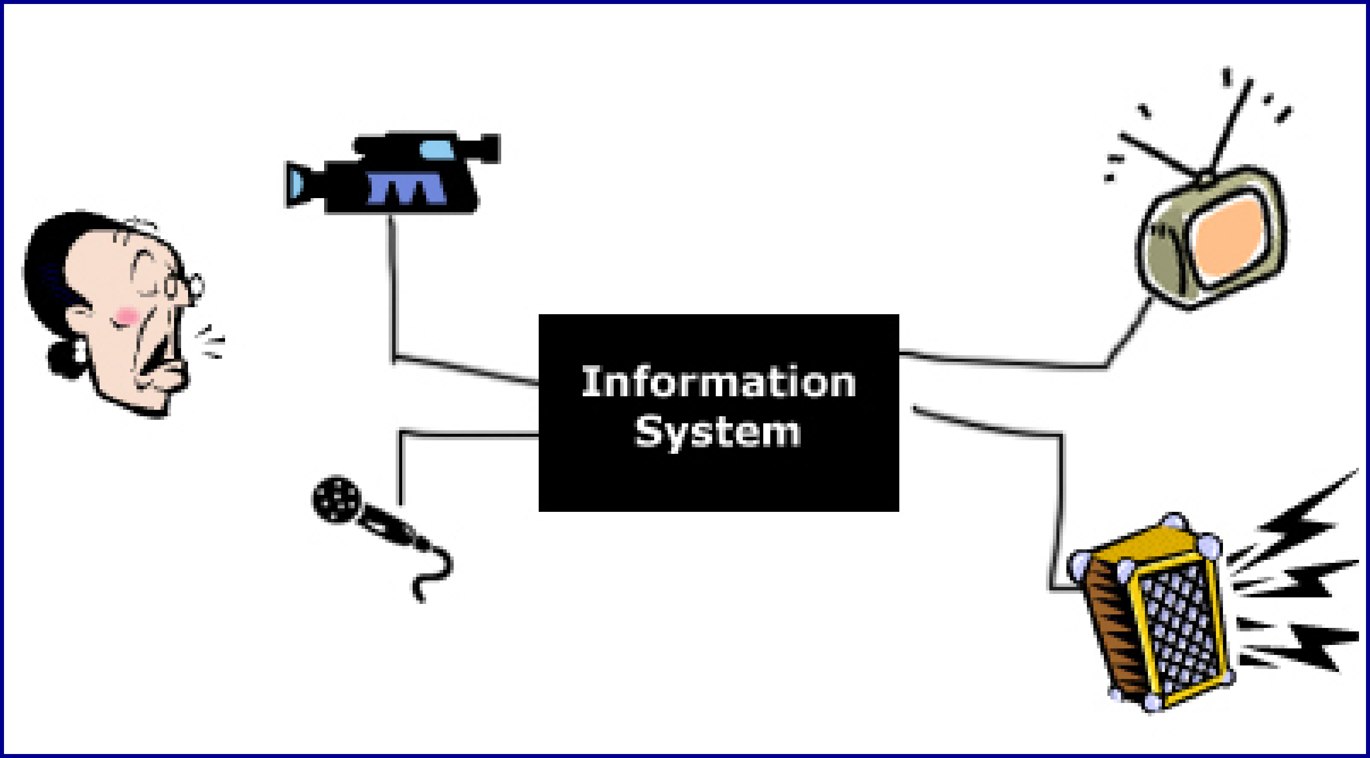
Roadmap

- Basic model
- Related tasks
- Face detection, facial gestures recognition
- Techniques
- Learning from examples

Roadmap

- Basic model
- Related tasks
- Face detection, facial gestures recognition
- Techniques
- Learning from examples
- Support vector machine and its application

Basic Model



Related Tasks

- Input:

Related Tasks

- Input: Speech recognition,

Related Tasks

- Input: Speech recognition, face detection,

Related Tasks

- Input: Speech recognition, face detection, facial gestures recognition,

Related Tasks

- Input: Speech recognition, face detection, facial gestures recognition, video-based speech recognition

Related Tasks

- Input: Speech recognition, face detection, facial gestures recognition, video-based speech recognition
- Engine:

Related Tasks

- Input: Speech recognition, face detection, facial gestures recognition, video-based speech recognition

- Engine: Knowledge-based solutions

Related Tasks

- Input: Speech recognition, face detection, facial gestures recognition, video-based speech recognition
- Engine: Knowledge-based solutions
- Output:

Related Tasks

- Input: Speech recognition, face detection, facial gestures recognition, video-based speech recognition
- Engine: Knowledge-based solutions
- Output: Speech synthesis,

Related Tasks

- Input: Speech recognition, face detection, facial gestures recognition, video-based speech recognition
- Engine: Knowledge-based solutions
- Output: Speech synthesis, talking-head

What is the Face Detection?

What is the Face Detection?

- Face detection

What is the Face Detection?

- Face detection
- Face tracking

What is the Face Detection?

- Face detection
- Face tracking
- Face recognition

What is the Face Detection?

- Face detection
- Face tracking
- Face recognition, face verification

What is the Face Detection?

- Face detection
- Face tracking
- Face recognition, face verification
- Facial gestures recognition

Detecting Faces in Still Images – Problems

- Pose: The images of a face vary due to the relative camera-face pose.

Detecting Faces in Still Images – Problems

- Pose: The images of a face vary due to the relative camera-face pose.
- Presence or absence of structural components: beards, mustaches, glasses.

Detecting Faces in Still Images – Problems

- Pose: The images of a face vary due to the relative camera-face pose.
- Presence or absence of structural components: beards, mustaches, glasses.
- Facial expressions

Detecting Faces in Still Images – Problems

- Pose: The images of a face vary due to the relative camera-face pose.
- Presence or absence of structural components: beards, mustaches, glasses.
- Facial expressions
- Occlusion: Faces may be partially occluded by other objects.

- Imaging conditions: Lighting and camera characteristics.

Detecting Faces in Still Images – Solutions

- Knowledge-based method:

Detecting Faces in Still Images – Solutions

- Knowledge-based method: Encode human knowledge of what constitutes a typical face.

Detecting Faces in Still Images – Solutions

- Knowledge-based method: Encode human knowledge of what constitutes a typical face.

- Feature invariant approaches:

Detecting Faces in Still Images – Solutions

- Knowledge-based method: Encode human knowledge of what constitutes a typical face.
- Feature invariant approaches: Aim to find structural features of a face that exist even when the pose,viewpoint, or lighting conditions vary.

- Template matching methods:

- Template matching methods: Several standard patterns stored to describe the face as a whole or the facial features separately.

- Template matching methods: Several standard patterns stored to describe the face as a whole or the facial features separately.

- Appearance-based methods:

- Template matching methods: Several standard patterns stored to describe the face as a whole or the facial features separately.

- Appearance-based methods: The models are learned from a set of training images which capture the representative variability of facial appearance.

Face Detection

Face Detection

- Equalization of the gray-level information

Face Detection

- Equalization of the gray-level information
- Oval mask

Face Detection

- Equalization of the gray-level information
- Oval mask
- Scanning

Face Detection

- Equalization of the gray-level information
- Oval mask
- Scanning
- Extraction of the information to a pyramid.

Face Detection

- Equalization of the gray-level information
- Oval mask
- Scanning
- Extraction of the information to a pyramid.
- SVM

Model of Learning from Examples

- A generator of random vectors x, drawn independtly from fixed, but unkown distribution $P(x)$.

Model of Learning from Examples

- A generator of random vectors x, drawn independtly from fixed, but unkown distribution $P(x)$.

- A supervisor that returns an output vector y for every input vector x, according to a conditional distribution function $P(y|x)$, also fixed but unkown.

Model of Learning from Examples

- A generator of random vectors x, drawn independtly from fixed, but unkown distribution $P(x)$.
- A supervisor that returns an output vector y for every input vector x, according to a conditional distribution function $P(y|x)$, also fixed but unkown.
- A learning machine capable of implementing a set of functions $f(x, \alpha)$, $\alpha \in \Lambda$.

- The problem of learning is that of choosing from the given set of function, the one which predicts the supervisor's response in the best possible way. The selection is based on a training set of l random independent identically distributed observations drawn according to $P(x, y) = P(x)P(y|x)$.

Problem of Risk Minimization

- In order to choose the best available approximation to the supervisor's response, one measures the loss $L(y, f(x, \alpha))$ between the response y of the supervisor to a given input x and the response $f(x, \alpha)$ provided by the learning machine. Consider the expected value of the loss, given by the risk functional

$$R(\alpha) = \int L(y, f(x, \alpha)) dP(x, y).$$

- The goal is to find the function $f(x, \alpha_0)$ which minimizes the risk functional $R(\alpha)$ in the situation where the joint probability distribution $P(x, y)$ is unknown and the only available information is contained in the training set.

The Optimal Separating Hyperplanes

- Suppose the training data

$$(x_1, y_1), \ldots, (x_l, y_l), x \in \mathbf{R}^n, y \in \{+1, -1\}$$

can be separated by a hyperplane

$$(w \cdot x) - b = 0.$$

- We say that this set of vectors is separated by the optimal hyperplane if it is separated without error and the distance between the closest vector and the hyperplane is maximal.

- We say that this set of vectors is separated by the optimal hyperplane if it is separated without error and the distance between the closest vector and the hyperplane is maximal.

- To describe the separating hyperplane let us use the following form:

$$(w \cdot x_i) - b \geq 1, \text{if } y_i = +1,$$

$$(w \cdot x_i) - b \leq 1, \text{if } y_i = -1.$$

- In the following we use a compact notation for these inequalities:

$$y_i((w \cdot x_i) - b) \geq 1, i = 1, \ldots, 1.$$

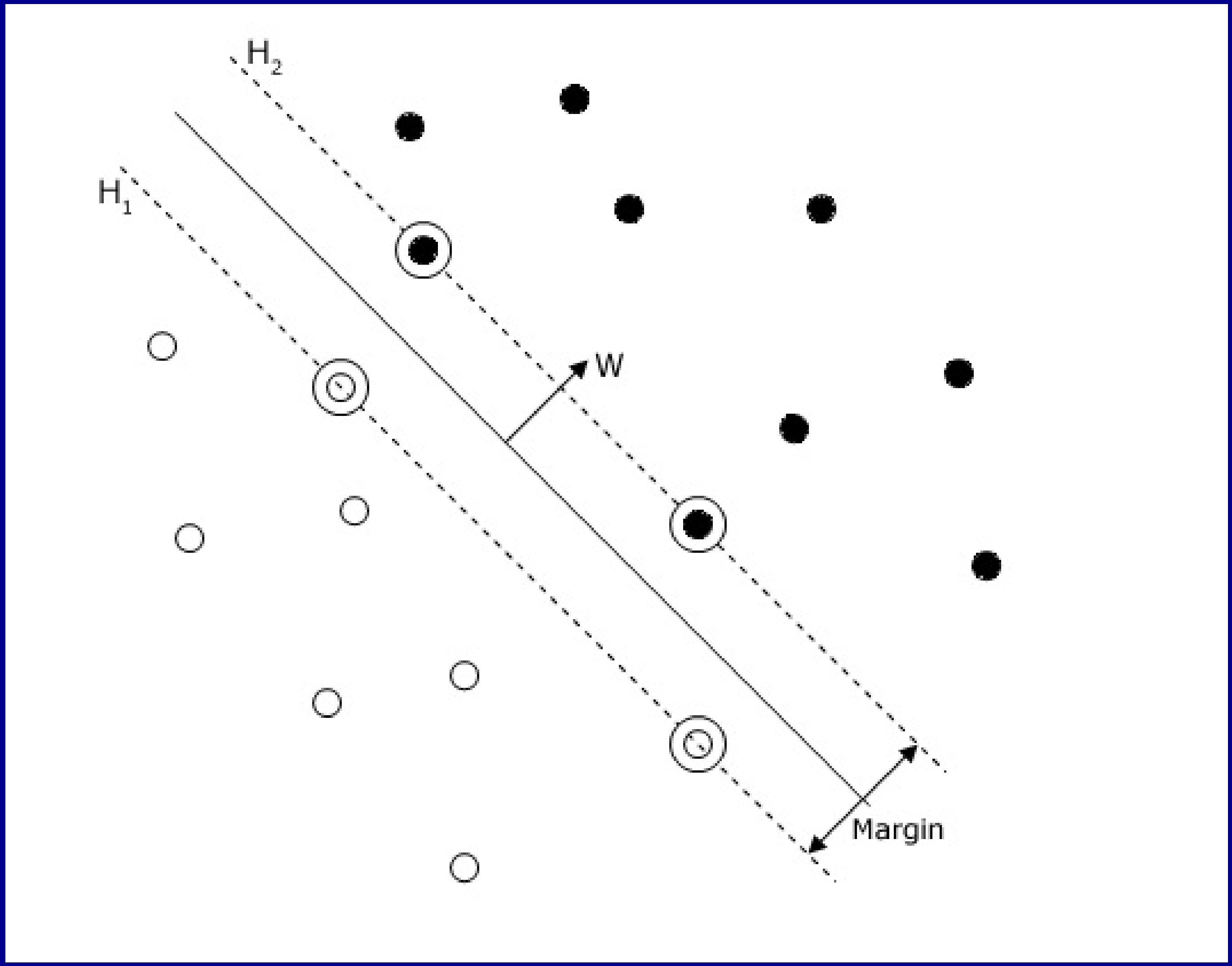
- In the following we use a compact notation for these inequalities:

$$y_i((w \cdot x_i) - b) \geq 1, i = 1, \ldots, 1.$$

- It is easy to check that the optimal hyperplane is the one that satisfies the condition and minimizes functional

$$\Phi(w) = \frac{1}{2} \cdot \|w\|^2.$$

- The solution to this optimization problem is given by the saddle point of a Lagrange functional.


H2
H1
W
Margin

The Problem of Pattern Recognition

- Let the supervisor's output y take on only two values $y = \{0, 1\}$.

The Problem of Pattern Recognition

- Let the supervisor's output y take on only two values $y = \{0, 1\}$.

- Let $f(x, \alpha)$, $\alpha \in \Lambda$ be a set of indicator functions (functions which take on only two values zero and one).

- Consider the following loss-function

$$L(y, f(x,\alpha)) = \begin{cases} 0 & \text{, if } y = f(x,\alpha), \\ 1 & \text{, if } y \neq f(x,\alpha). \end{cases}$$

- Consider the following loss-function

$$L(y, f(x,\alpha)) = \begin{cases} 0 & \text{, if } y = f(x,\alpha), \\ 1 & \text{, if } y \neq f(x,\alpha). \end{cases}$$

- The poblem is to find the function which minimizes the probability of classification errors when probability measure $P(x, y)$ is unkown, but the data are given.

The Importance of the Set of Functions

- What about allowing all functions from $\mathbf{R}^N$ to $\{\pm 1\}$?

The Importance of the Set of Functions

- What about allowing all functions from $\mathbf{R}^N$ to $\{\pm 1\}$?
- Training set $(x_1, y_1), \ldots, (x_l, y_l) \in \mathbf{R}^N \times \{\pm 1\}$.

The Importance of the Set of Functions

- What about allowing all functions from $\mathbf{R}^N$ to $\{\pm 1\}$?

- Training set $(x_1, y_1), \ldots, (x_l, y_l) \in \mathbf{R}^N \times \{\pm 1\}$.

- Test patterns $\overline{x_1}, \ldots, \overline{x_l} \in \mathbf{R}^N$, such that the elements of the training set is not elements of the test set.

- Based on the training set alone, there is no means of choosing which one is better, because for any f there exits f^*, where
  - $f^*(x_i) = f(x)$, for all i,
  - $f^*(\overline{x}_j \neq f(\overline{x}_j)$, for all j.

- Based on the training set alone, there is no means of choosing which one is better, because for any f there exits f^*, where
  - $f^*(x_i) = f(x)$, for all i,
  - $f^*(\overline{x}_j \neq f(\overline{x}_j)$, for all j.

- There is "no free lunch". The restriction must be placed on the functions that we allow.

Basic Example

- The problem we look at initially is the problem of finding binary classifiers.

Basic Example

- The problem we look at initially is the problem of finding binary classifiers.
- Let us consider the given weight and height of a person. We want to find a way of determining their gender.

- If we are given a set of examples with height, weight and gender, we can come up with a hypothesis which will enable us to determine a person's gender from their weight and height.

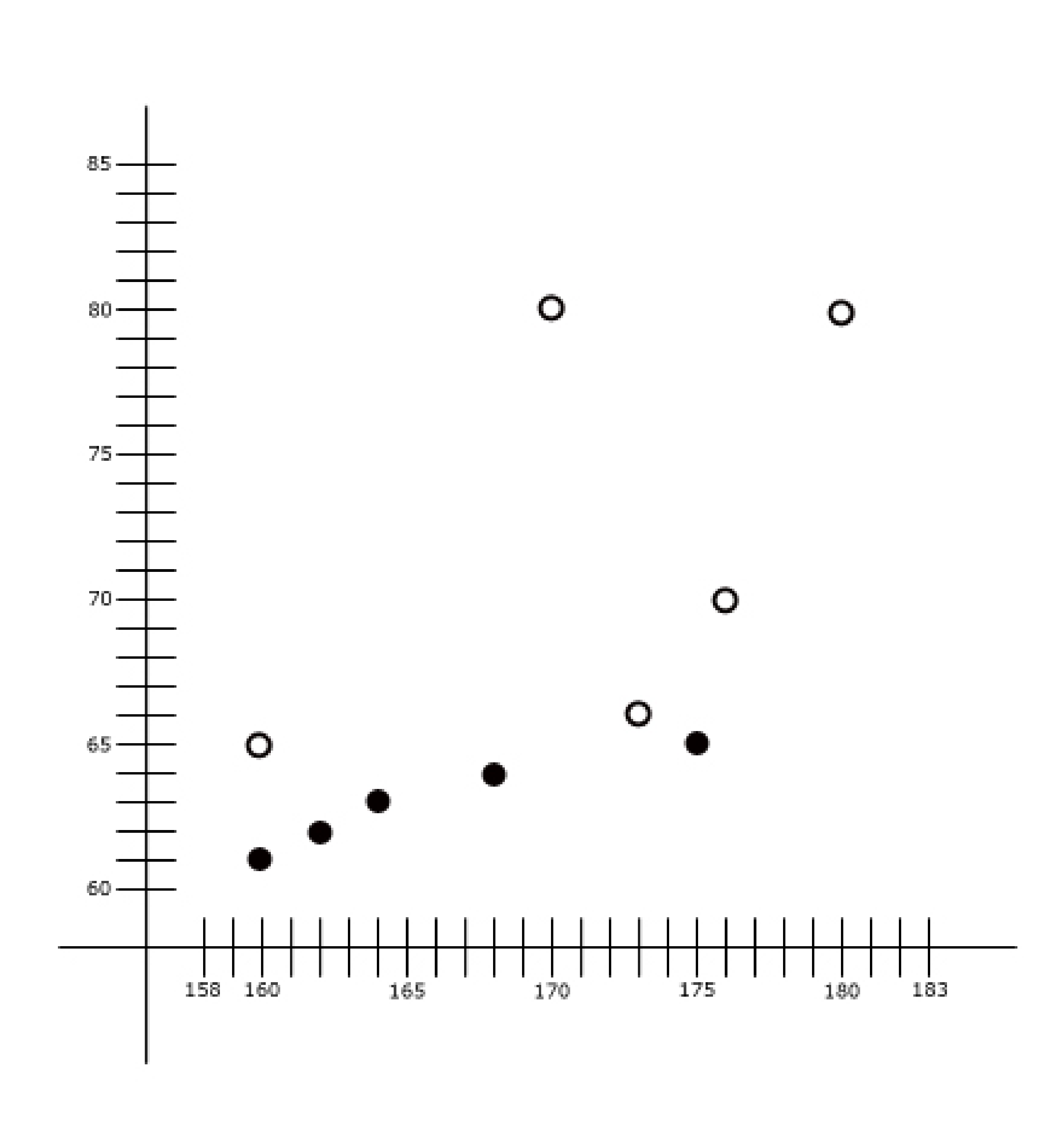
- If we are given a set of examples with height, weight and gender, we can come up with a hypothesis which will enable us to determine a person's gender from their weight and height.
- The weights and heights in a two-dimensional coordinate system are points.

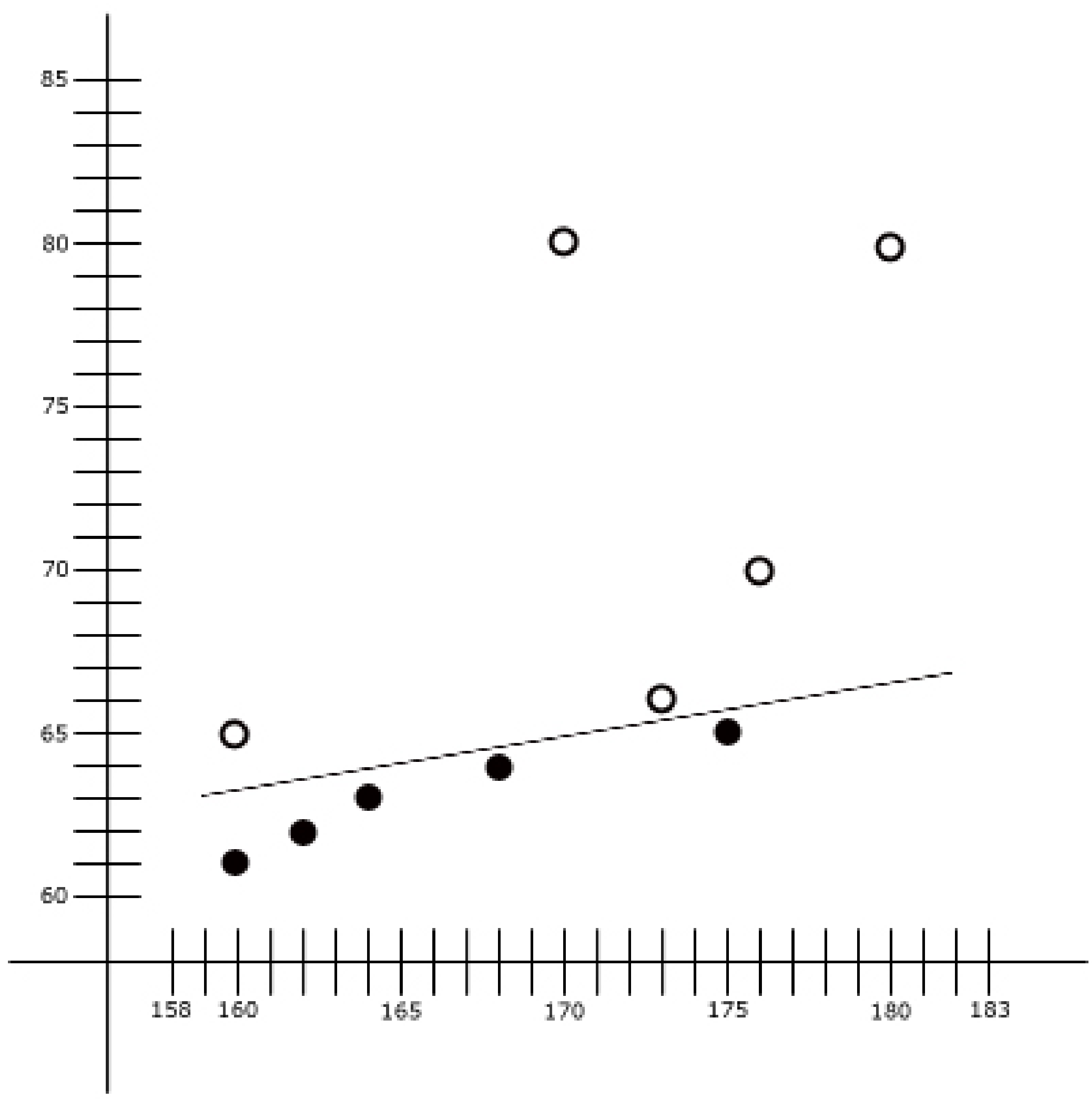
- If we are given a set of examples with height, weight and gender, we can come up with a hypothesis which will enable us to determine a person's gender from their weight and height.

- The weights and heights in a two-dimensional coordinate system are points.

- Let us find the separating hyperplane which divides the points into two regions, one female, one male.

No.	Height	Weight	Gender
1	180	80	m
2	173	66	m
3	170	80	m
4	176	70	m
5	160	65	m
6	160	61	f
7	162	62	f
8	168	64	f
9	164	63	f
10	175	65	f


85
80
75
70
65
60
158
160
165
170
175
180
183


85
80
75
70
65
60
158
160
165
170
175
180
183

About the VC-dimension

- The Vapnik-Chervonenkis dimension has a very important role in the statistical learning.

About the VC-dimension

- The Vapnik-Chervonenkis dimension has a very important role in the statistical learning.
  - It characterizes the learning capacity.

About the VC-dimension

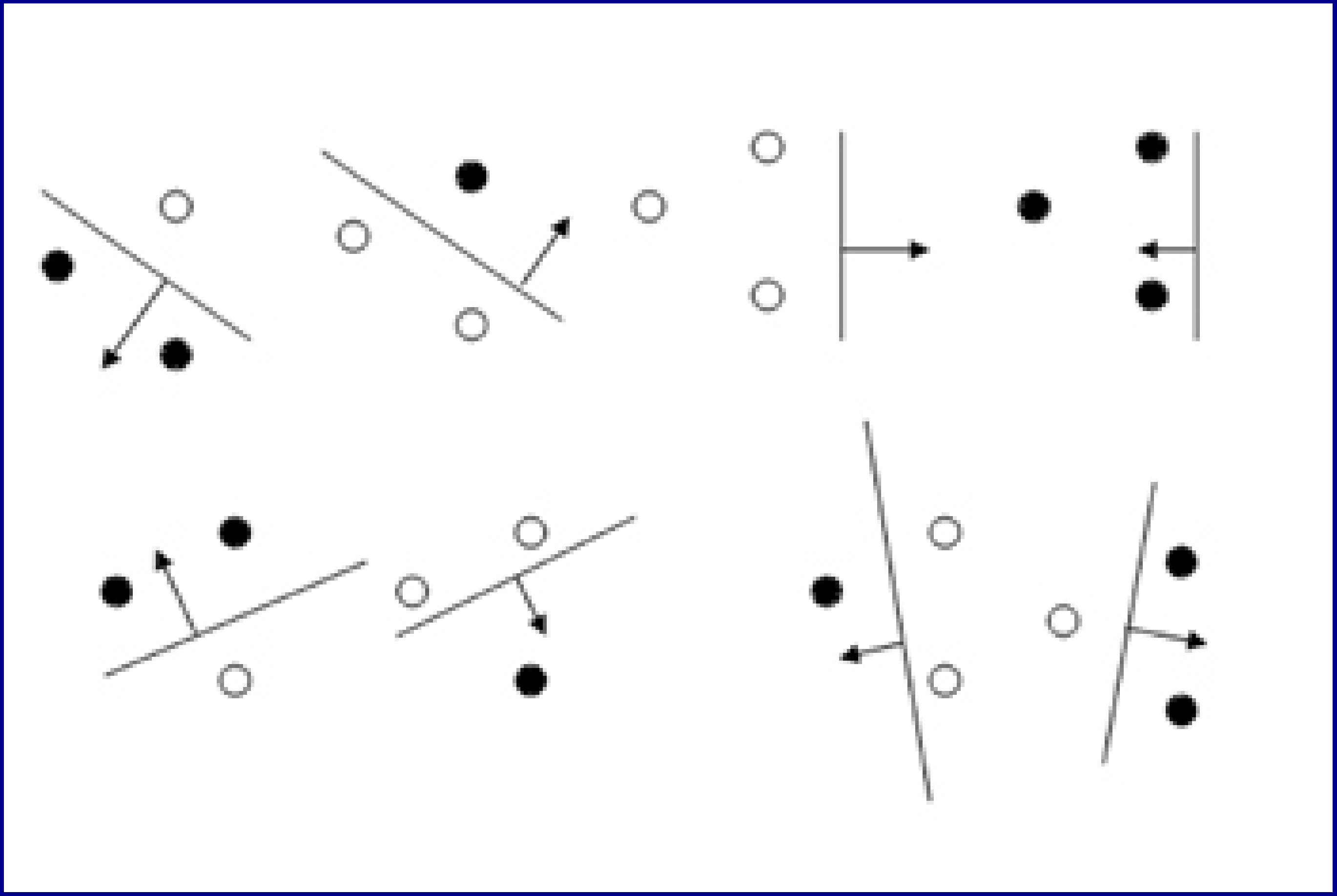
- The Vapnik-Chervonenkis dimension has a very important role in the statistical learning.
  - It characterizes the learning capacity.
  - One can avoid the overfitting with its control.

About the VC-dimension

- The Vapnik-Chervonenkis dimension has a very important role in the statistical learning.
  - It characterizes the learning capacity.
  - One can avoid the overfitting with its control.
  - One can minimize the expected value of the error with its control.

- The VC-dimension of a set of $+1, -1$-valued functions is equal to the largest number h of points of the domain of the functions that can be separated into two different classes in all the 2^h possible ways using the functions of this set of functions.

Determine the VC-dimension!



Support Vector Machine

- Map the input vectors into a very high-dimensional feature space through some nonlinear mapping choosen a priori.

Support Vector Machine

- Map the input vectors into a very high-dimensional feature space through some nonlinear mapping choosen a priori.
- In this space construct an optimal separating hyperplane.

Support Vector Machine

- Map the input vectors into a very high-dimensional feature space through some nonlinear mapping choosen a priori.
- In this space construct an optimal separating hyperplane.
- To generalize well, we control (decrease) the VC dimension by constructing an optimal separating hyperplane (that maximizes the margin).

- To increase the margin we use very high dimensional spaces.

- To increase the margin we use very high dimensional spaces.

- The training algorithm would only depend on the data through dot products in the feature space, i.e. on functions of the form $\Phi(x_i) \cdot \Phi(x_j)$. Now if there were a "kernel function" K such that $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$, we would only need to use K in the training algorithm, and would never need to explicitly even know what Φ is.

- Polynomial kernel $K(x, y) = (x \cdot y + 1)^p$

- Polynomial kernel $K(x, y) = (x \cdot y + 1)^p$
- Gaussian radial kernel $K(x, y) = e^{-\|x-y\|^2/2\sigma^2}$

- Polynomial kernel $K(x,y) = (x \cdot y + 1)^p$
- Gaussian radial kernel $K(x,y) = e^{-\|x-y\|^2/2\sigma^2}$
- Two-layer sigmoidal neural network $(x,y) = \tanh(\kappa x \cdot y - \delta)$

Summary of Some Features of SVM

- It creates a classifier with minimised VC-dimension.

Summary of Some Features of SVM

- It creates a classifier with minimised VC-dimension.
- If the VC dimension is low, the expected probability of error is low as well.

- SVM uses a linear separating hyperplane to create a classifier. But some problems can not be linearly separated in the original input space.

- SVM uses a linear separating hyperplane to create a classifier. But some problems can not be linearly separated in the original input space.

- SVM can non-linearly transform the original input space into a higher dimensional feature space.

Experimental Results

- For all experiments the Mathlab SVM toolbox developed by Steve Gunn was used. For a complete test, several auxiliary routines have been added to the original toolbox.

- Images

- Images
  - Training set of 46 images (31 face, 15 non-face)

- Images
  - Training set of 46 images (31 face, 15 non-face)
  - Ibermatica – several sources of degradation are modeled.

- Images
  - Training set of 46 images (31 face, 15 non-face)
  - Ibermatica – several sources of degradation are modeled.
  - All images are recorded in 256 grey levels.

- Images
  - Training set of 46 images (31 face, 15 non-face)
  - Ibermatica – several sources of degradation are modeled.
  - All images are recorded in 256 grey levels.
  - They are of dimension 320×240.

- The procedure for collecting face patterns is as follows.

- The procedure for collecting face patterns is as follows.
  - A rectangle part of dimensions 128×128 pixels has been manually determined that includes the actual face.

- The procedure for collecting face patterns is as follows.
  - A rectangle part of dimensions 128×128 pixels has been manually determined that includes the actual face.
  - This area has been subsampled four times. At each subsampled four times. At each subsampling, non- overlapping regions of 2×2 are replaced by their average.

- The training patterns of dimension 8×8 are built.

- The training patterns of dimension 8×8 are built.
- The class label $+1$ has been appended to each pattern.

- The training patterns of dimension 8×8 are built.
- The class label $+1$ has been appended to each pattern.
- Similarly, 15 non-face patterns have been collected from images in the same way, and labeled by -1.

- We have trained the three different SVMs. The trained SVMs have been applied to 414 test examples (249 face and 165 non-face). The test images are classified as non-face ones or face ones. The following table gives the results on the test.

- We have trained the three different SVMs. The trained SVMs have been applied to 414 test examples (249 face and 165 non-face). The test images are classified as non-face ones or face ones. The following table gives the results on the test.

	Linear	Walsh	Polynomial
Time	2.3581	2.3432	2.5327
Errors	9	8	7
Margin	0.66	4.58	2.17
SVs	15	12	8