

# Introduction

Machine Learning Course

P. Viswanath

IIITS

# Overview

- History of AI
  - past, present, future.
- Techniques
  - Supervised learning techniques
  - Feature extraction
  -
- Some Challenges

# History of Technology

- 1771 Industrial Revolution

- Arkwright's mill in Cramford



- 1826 Age of Steam and Railways

- “Rocket steam” engine for Manchester railway



- 1875 Age of Steel, Electricity & Heavy Eng.

- Carnegie Bessemer steel plant in Pittsburgh



- 1908 Age of Oil, Automobile, Mass Production

- First model T comes out in Detroit



- 1971 Age of Information & Telecommunications

- Intel microprocessor announced in Santa Clara



- 2017 Age of Artificial Intelligence

- Machines, data and people connected in new era



- The term **artificial intelligence** was first coined in 1956, at the Dartmouth conference.
- Artificial Intelligence is still a growing active field of science and technology.
  - Has the potential to affect our lives

# AI?

- *Homo sapiens (human beings) are able to control (and exploit) other species and nature because of their thinking capability.*

We call programs intelligent if they exhibit behaviors that would be regarded intelligent if they were exhibited by human beings.

– Herbert Simon

# Turing Test

## Alan Turing's Imitation Game

Alan Turing (1912 – 1954)

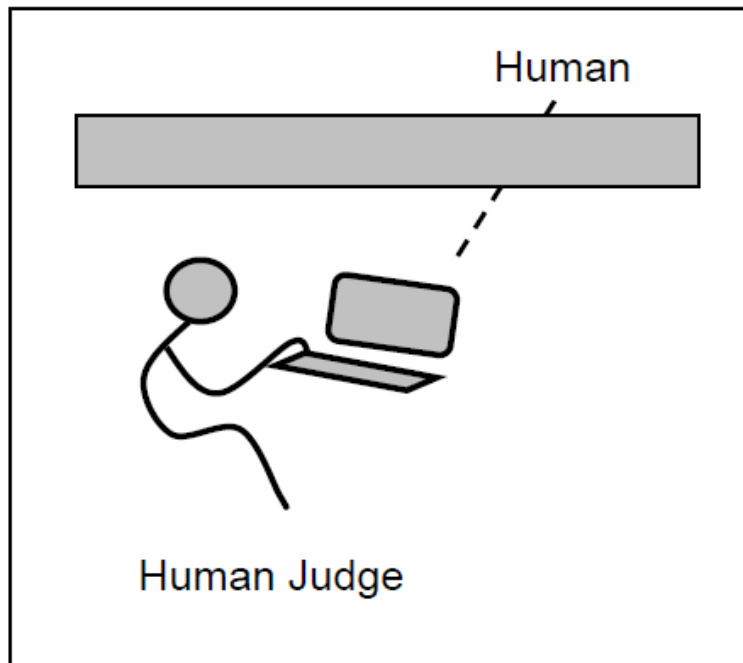
- The question whether machines can think itself “too meaningless”
- Prescribed a test which he called the *Imitation Game* which is now known as *The Turing Test*



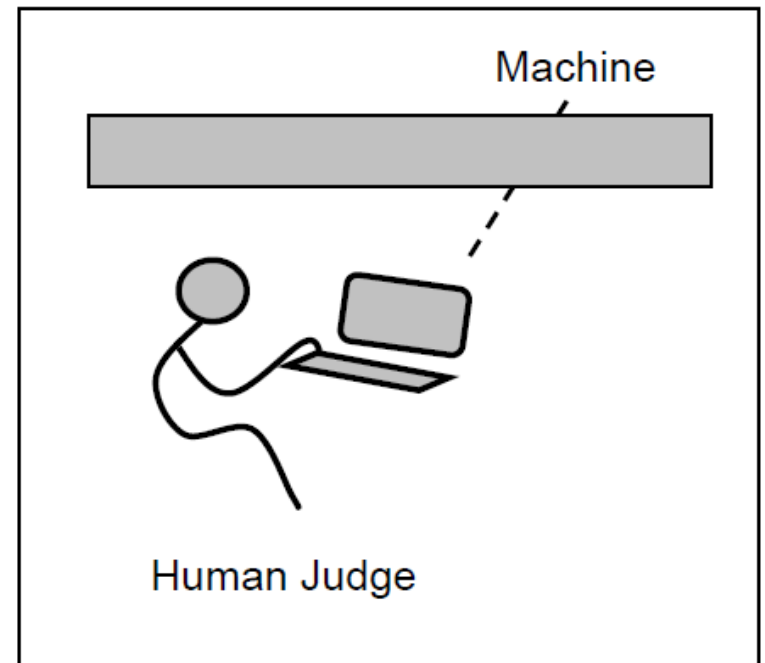
[http://en.wikipedia.org/wiki/Alan\\_Turing](http://en.wikipedia.org/wiki/Alan_Turing)

# Human or Machine?

## The Turing Test



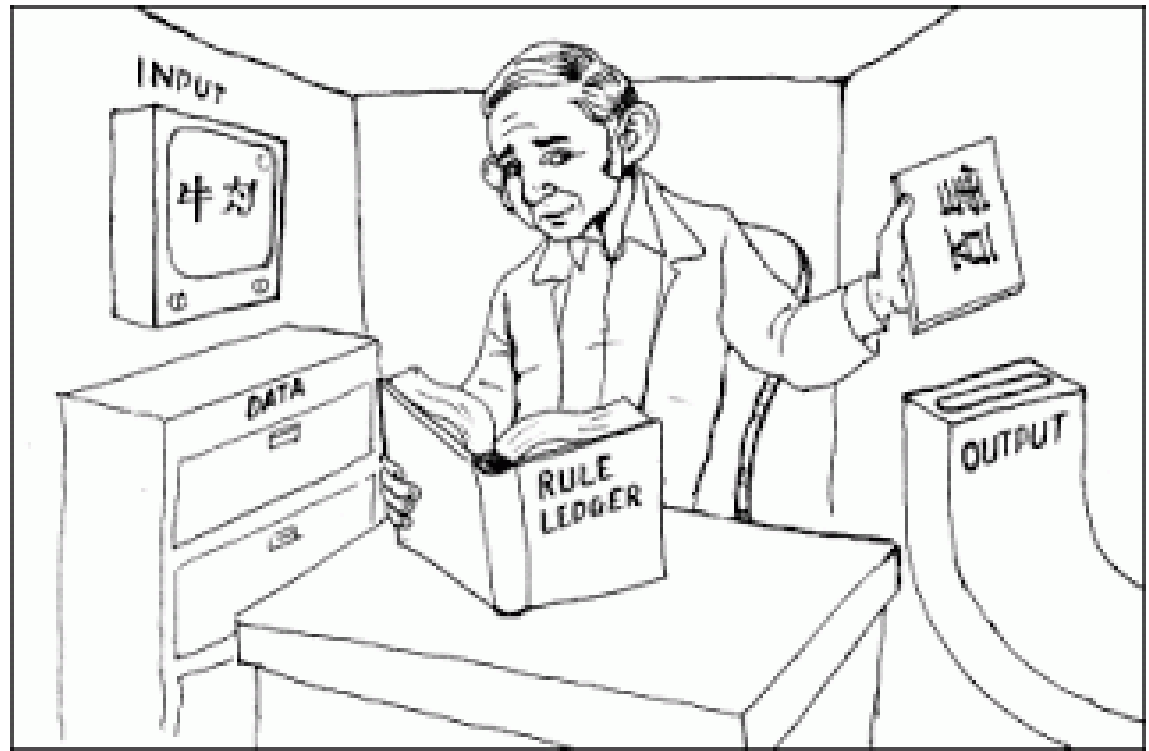
or?



The Loebner Prize – an annual competition where **chatbots** are judged for human like response. The grand prize of USD 100,000 is still open.

# The Chinese Room Argument

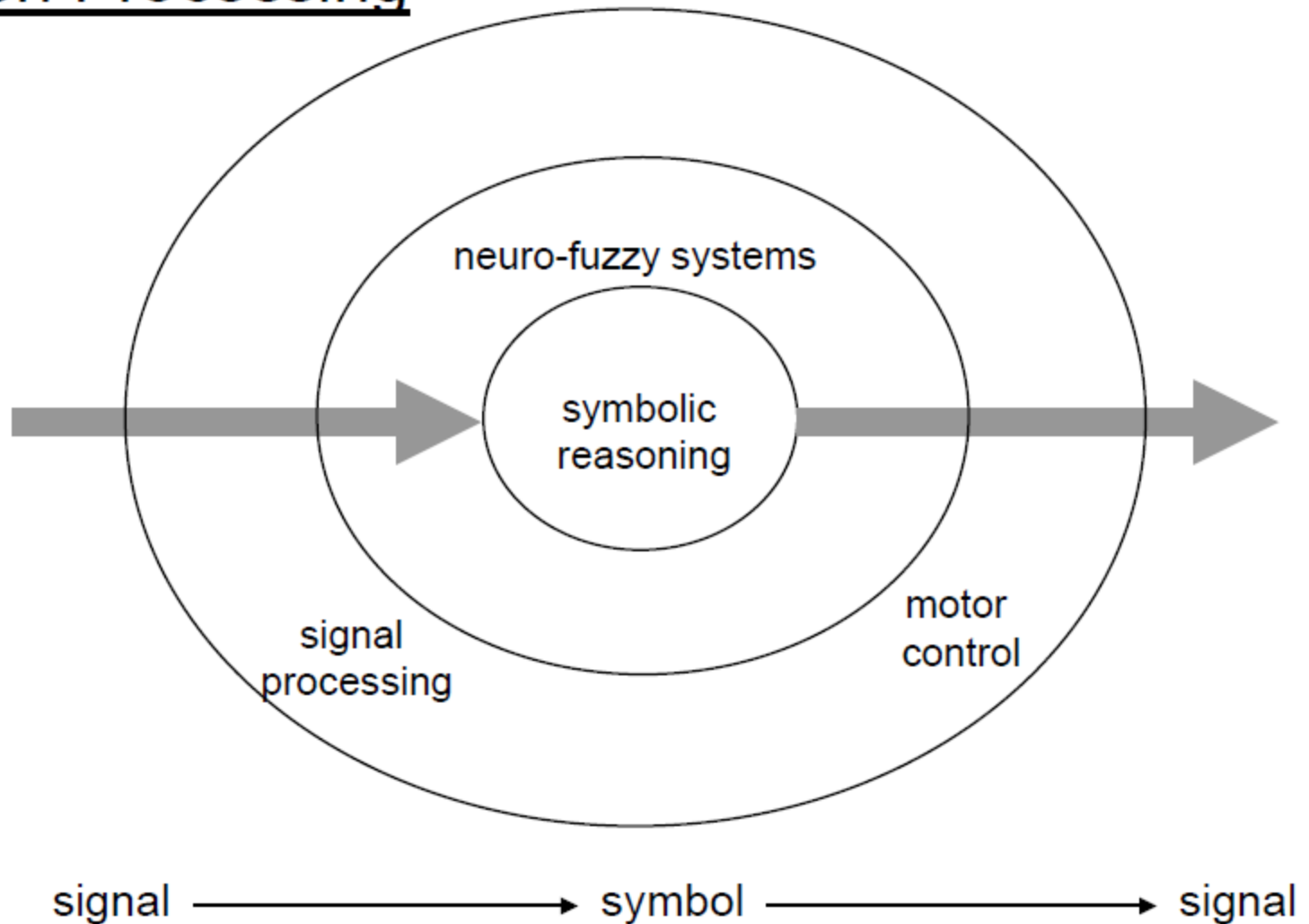
- John Searle: The Chinese Room argument – can an agent locked in a room processing questions in Chinese based on a set of syntactic rules be said to *understand* Chinese?
  - How many rules will the agent need to have for the thought experiment to be convincing?





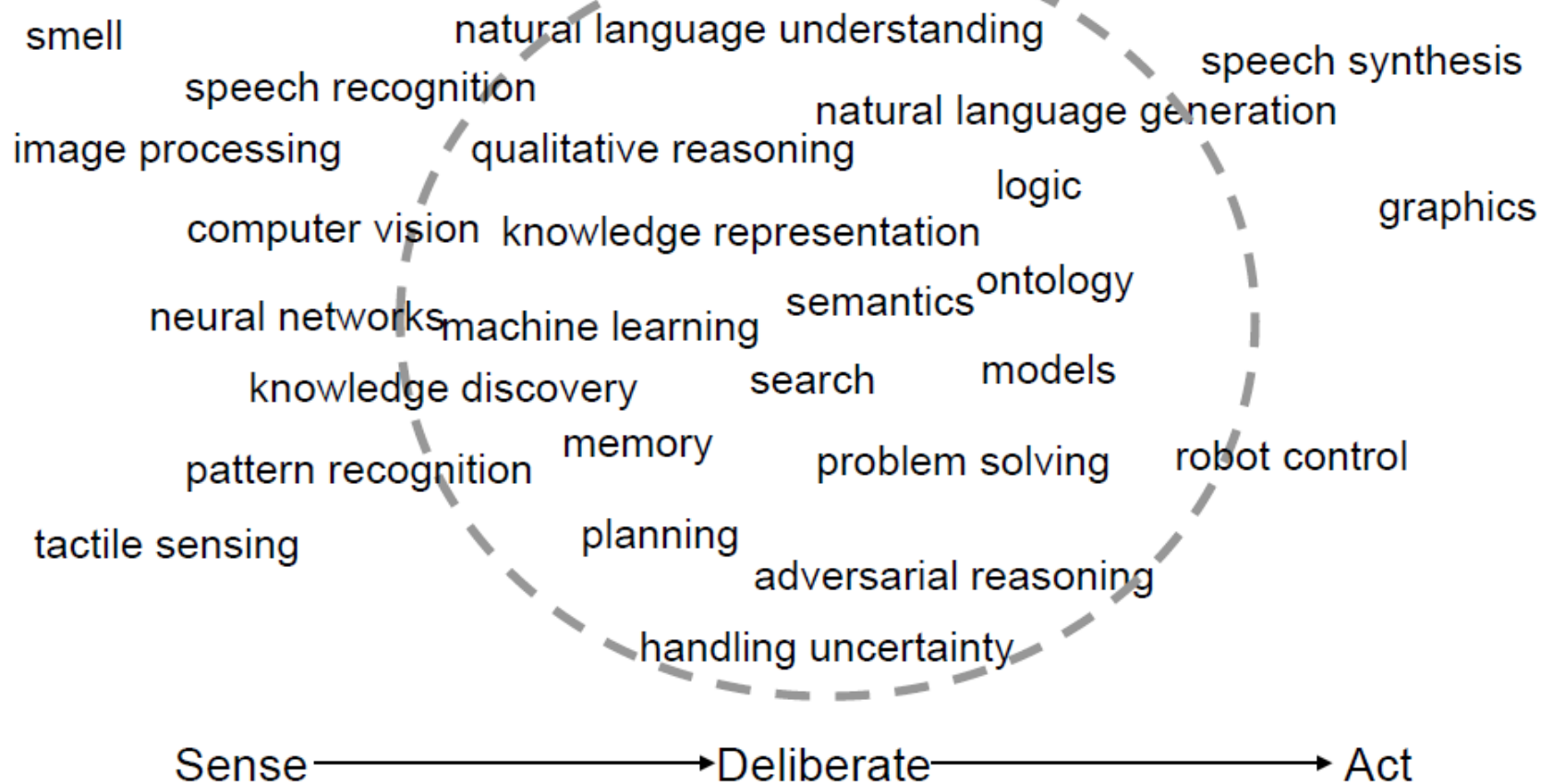
# Information Processing?

## Information Processing



# Topics in AI

## Topics in AI



# What we already achieved in AI?

- Board games – Chess, Checkers, Go, etc
- Solving Puzzles – Sudoku, etc
- Route finding in a map
- Image/speech enhancement
  - Creating high resolution images, noise suppression, ...
-

# Today's AI

- 
- A diagram of a neural network. On the left, there are two large blue arrows pointing right, representing input. These connect to a column of three blue circles representing the input layer. These circles are connected to a central column of three blue circles representing the hidden layer. The hidden layer circles are connected to a column of three blue circles on the right representing the output layer. There are also feedback connections from the output layer back to the hidden layer and from the hidden layer back to the input layer. The entire network is enclosed in a green box.

# Future prospects of AI

## Survey of AI researchers

### – AI will outperform humans in:

- Translating languages 2024
- Writing high-school essays 2026
- Driving a truck 2027
- Working in retail 2031
- Writing a best-selling book 2049
- Working as a surgeon 2053
- Outperform humans in all tasks: 50% chance in 45 years
- Automating all human jobs 120 years

### – Survey population: 2015 NIPS/ICML authors

- Questions on AI capabilities (e.g. folding laundry, language translation), superiority at specific occupations (e.g. truck driver, surgeon), superiority over humans at all tasks.

# Machine Learning?

- ML is a subset of AI.
- We focus on mathematical/algorithmic aspects of learning which can be programmed over a machine.
- We look into various learning paradigms.
- Prerequisites of the course are
  - Probability theory
  - Linear algebra
  - Calculus

# Evaluation/Examinations

- 20 Marks (16.6%) for Quizzes (best 2 out of 3 taken)
- 25 Marks (20.8%) for Assignments (3 assignments)
- 15 Marks (12.5%) for Mid1
- 25 Marks (20.8%) for Mid2
- 35 Marks (29.16%) for Endsem.
- Total – 120 Marks (100%).

# Text books/ References

1. **“Pattern Classification”** by R. O. Duda, P. E. Hart and D. G. Stork.
2. **“An Introduction to Statistical Learning”** by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani.
3. **“Pattern Recognition and Machine Learning”** by Christopher M. Bishop.
4. **“Introduction to Machine Learning”** by Ethem Alpaydin.
5. **“Pattern Recognition: An Algorithmic Approach”** by M. Narasimha Murty, V. Susheela Devi.
6. **“Machine learning”** by Tom Mitchell.



# Programming Assignments

- We follow C and or Python.
- No other languages are allowed.
- Plagiarism verification is done.
  - Simply changing variable names may not work for you?!

# Examinations/Quizzes

- Small numeric problems are asked.
- You need to employ the techniques discussed in the class.
- Since, these are small numeric problems, just by trial and error you can get the answer.
  - But, this is not respected.
  - You must follow the method described in the class.

**LET US BEGIN ...**

# Deductive Vs Inductive Learning

- Deductive
  - Rules of the game are (hard coded) given ahead.
  - Eg: An algorithm to do multiplication of numbers is given. Given any two numbers you can apply this and get the answer.
- Inductive
  - We are given with examples (not the concept). We need to learn the mapping from i/p to o/p.
    - Supervised learning problems in AI comes under this

# Learning strategies

- Supervised
  - Classification, Regression, ...
- Unsupervised
  - Clustering, density estimation, ...
- Reinforced
  - A robot navigating through obstacles, ...
- Learn the good features (attributes)
  - Feature extraction

# Spectrum of supervision

Less

More



Unsupervised



"Semi" supervised



Fully supervised

# What is a classification problem?

- Let there are two classes of objects.
  - Class 1: Set of dog pictures
  - Class 2: Set of cat pictures
- Problem is –
  - Given a picture, you should say whether it is cat or dog.
  - For a human being it is easy..., but for a machine it is a non-trivial problem.

# Training (Learning phase)

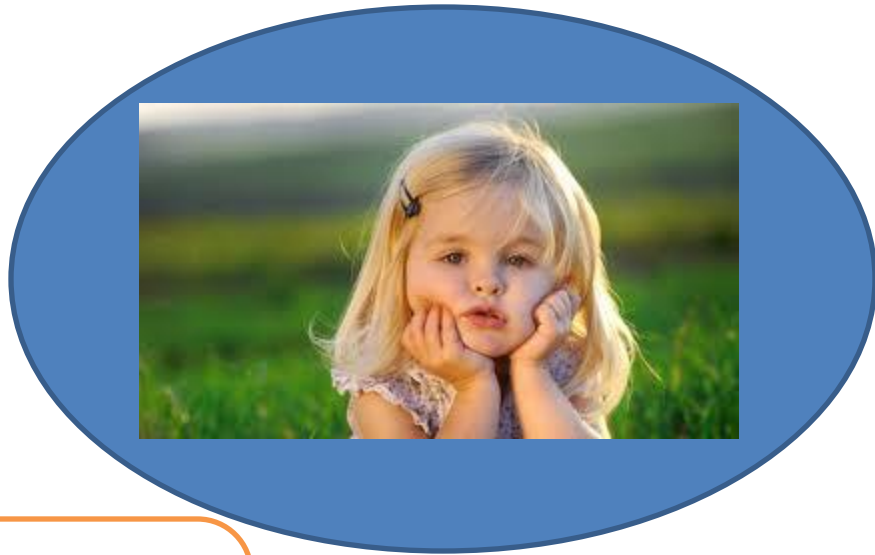


We have shown a set of dog pictures and a set of cat pictures to a child.





# Testing phase



DOG

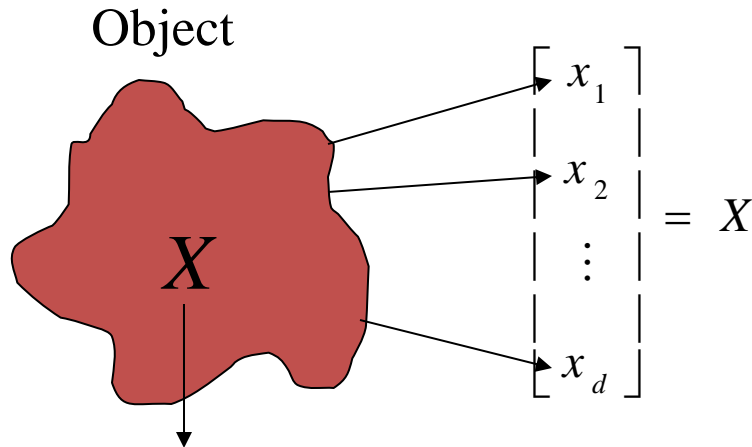
This picture as it is  
may not be in the  
training set

**Child has done more than  
just remembering**

# What is learning (pattern recognition)?

- Child has learnt what is it that is common among dogs ... and, what is it that is common among cats... also, what are the distinguishing features/attributes.
- Child has learnt the pattern (regularity) behind all dogs and the pattern behind all cats.
- Child then recognized a test image as having a particular pattern that is unique to dogs.

# Basic concepts



Feature vector  $X \in \mathcal{X}$

- A vector of observations (measurements).
- $X$  is a point in feature space  $\mathcal{X}$ .

Class to which  $X$  belongs is  $y \in Y$

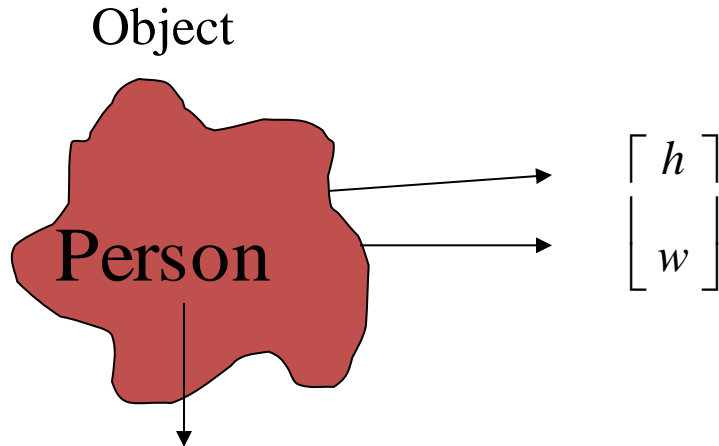
-Needs to be estimated, based on training set.

## Task

- To design a classifier (decision rule)  $f : \mathcal{X} \rightarrow Y$  which decides about the class label based on  $X$ .

# An example

$\mathcal{X}$  is a set of persons



## Feature vector

- A vector of observations (height, weight).

## Class to which $X$ belongs is

$y \in \{\text{overweight}, \text{normal}\}$

-Needs to be estimated, based on training set.

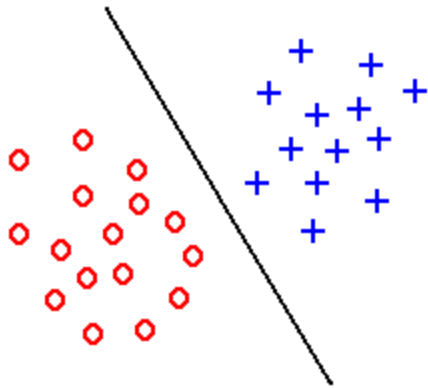
## Task

- To design a classifier (decision rule)  $f : \mathcal{X} \rightarrow Y$
- given height and weight of a person, classify him/her.

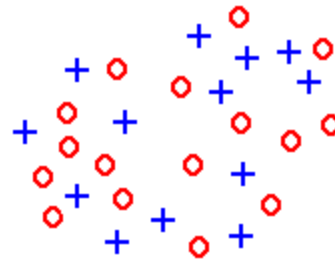
# Feature extraction

Task: to extract features which are good for classification.

Good features: • Objects from the same class have similar feature values.  
• Objects from different classes have different values.

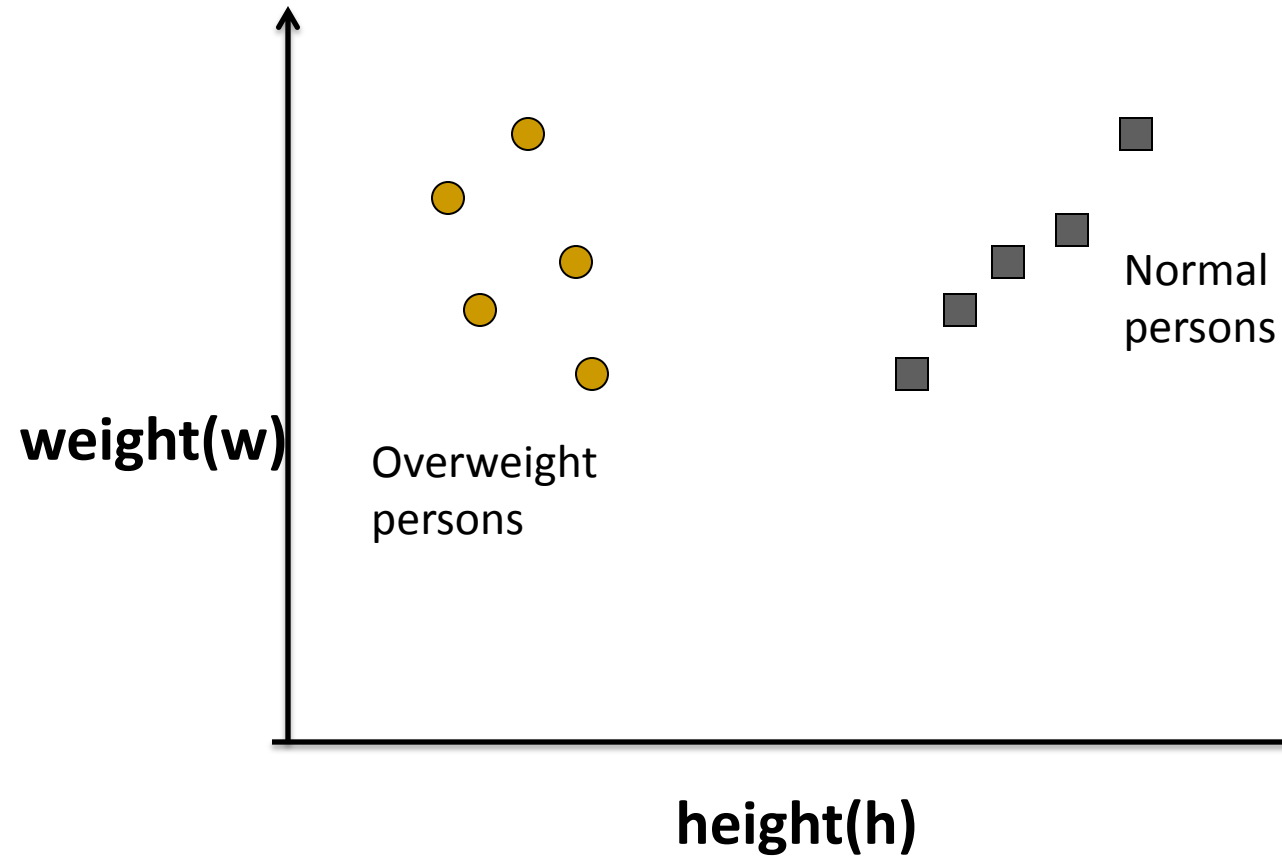


“Good” features



“Bad” features

# Feature Space



**Training set is shown in the feature space**

# Learning Steps

- Feature extraction: This is an important step. Good features are needed.
  - This is a lower level step. Normally done by techniques like image processing, speech processing, video processing, etc.
- Training set: Set of feature vectors along with their class labels.
  - An expert can see a few examples and give labels to them based on his experience.
- Build the classifier by using the training set.

# Classification Problem

- Given a training set, build the classifier.
- One has to evaluate, how good is the built classifier.
  - Of course, it has to agree with the training set
    - Is this 100% true?
  - But, it should do more than this.
  - The behavior of the classifier when it is asked to classify some thing which is not in the training set determines the quality.



# An easy, but bad classifier

- Remember the training set.
- See whether the given feature vector to be classified is available in the training set.
- If yes, then return the label of that training example.
- Else return a random class label.
- ***This is called Rote learning***

# Classifiers

- There are many classification methods.
  - Baye's classifier, Naïve Bayes classifier
  - HMM (graphical model)
  - Artificial Neural Networks
  - Decision Trees
  - SVMs
  - ....

# Generative vs. Discriminative Classifiers

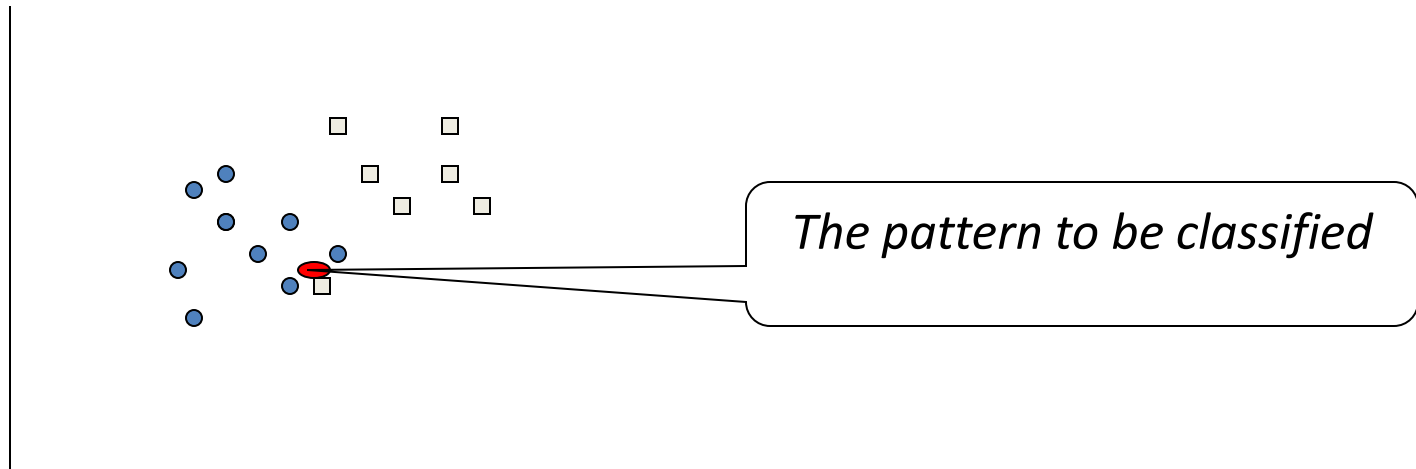
## Generative Models

- Learn the source. May be distribution is learnt.
- Often, makes use of conditional independence and priors
- Examples
  - Naïve Bayes classifier
  - Bayesian network
- Can be used to generate new examples !

## Discriminative Models

- Learn to directly predict the labels from the data
- Often, assume a simple boundary (e.g., linear)
- Examples
  - Logistic regression
  - SVM
  - Boosted decision trees
- Doesn't have the capacity to generate new examples.

# k-Nearest Neighbor Classifier



If  $k = 1$  then the class assigned is  $\square$

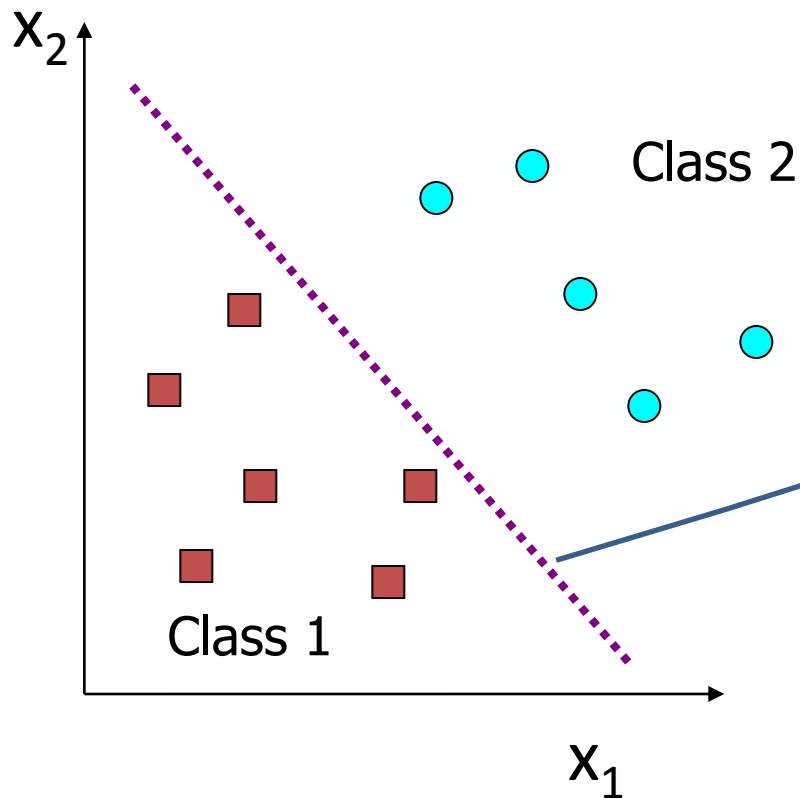
If  $k = 3$  then the class assigned is  $\bullet$

# Linear Classifier

## Classifier:

If  $f(x_1, x_2) < 0$  assign Class 1;

If  $f(x_1, x_2) > 0$  assign Class 2;



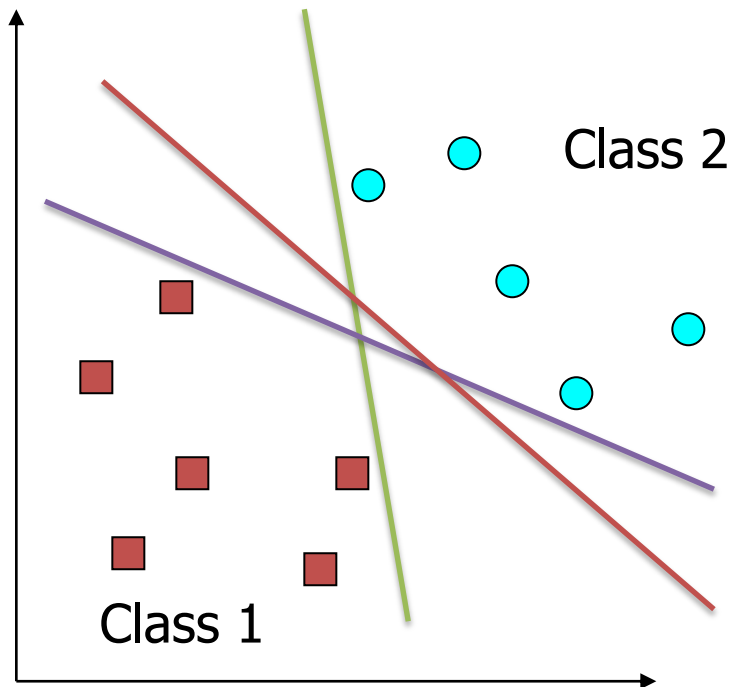
$$f(x_1, x_2) = w_1 x_1 + w_2 x_2 + b = 0$$

# Perceptron

- Perceptron is the name given to the linear classifier.
- If there exists a Perceptron that correctly classifies all training examples, then we say that the training set is **linearly separable**.
- In 1960s Rosenblatt gave an algorithm for Perceptron learning for linearly separable data.

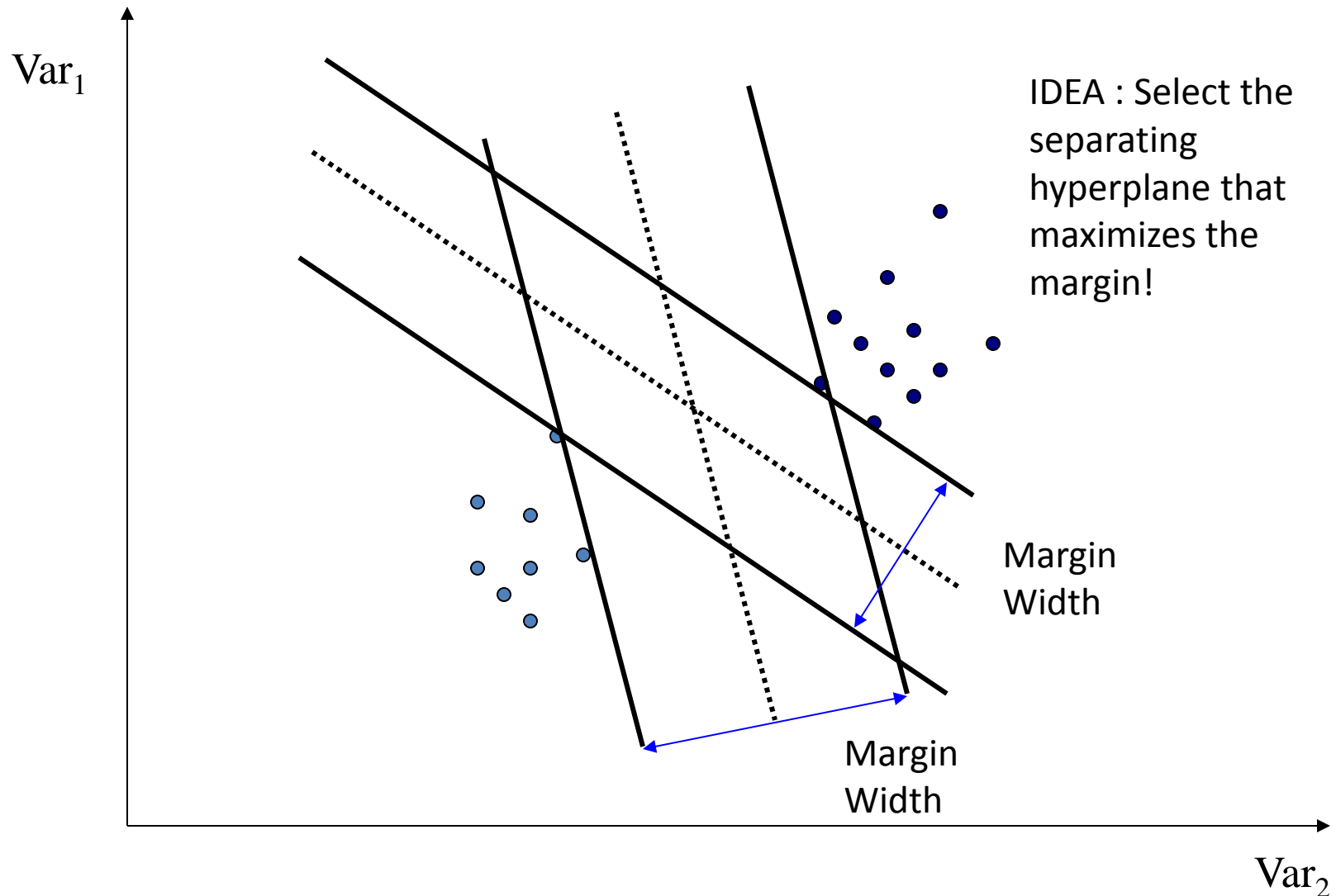
# Perceptron

- For linearly separable data many classifiers are possible.



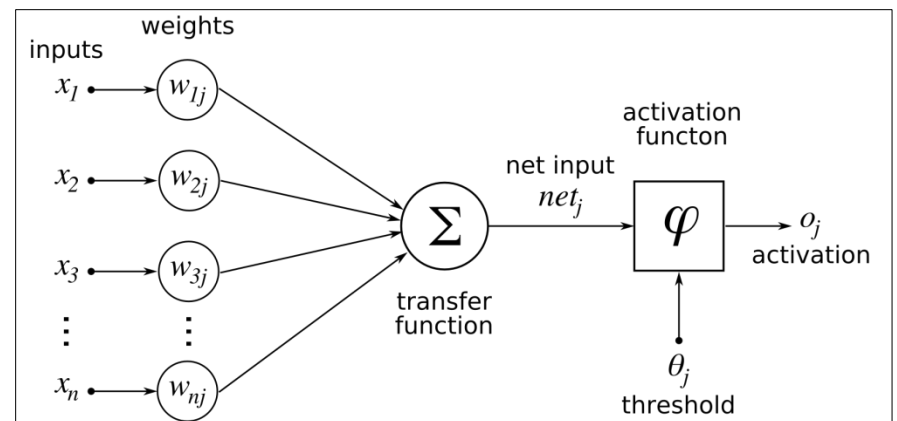
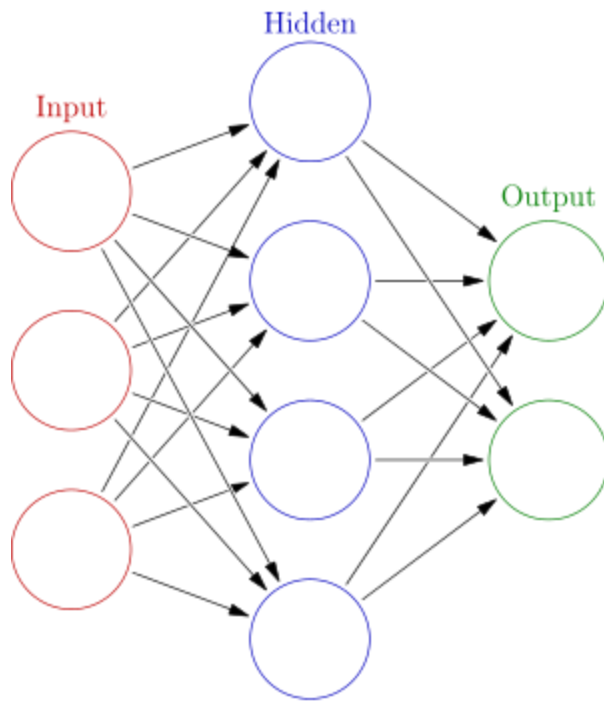
All being doing equally good on training set, which one is good on the unseen test set?

# Maximizing the Margin → SVM



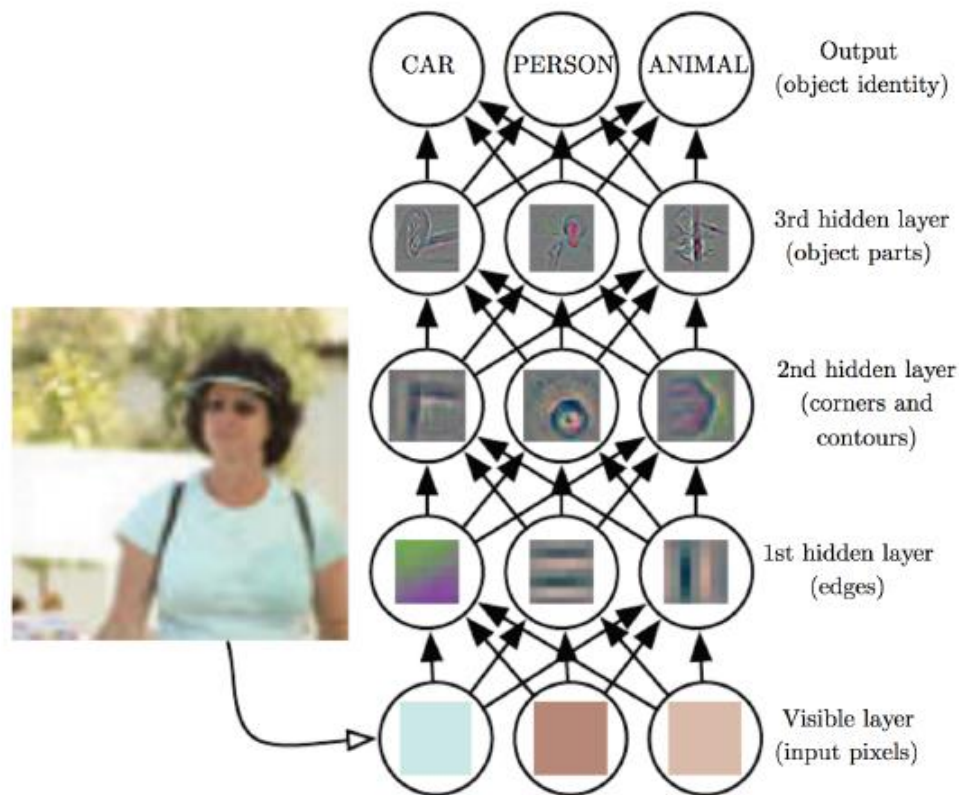


# Artificial Neural Networks



# Illustration of Deep Learning

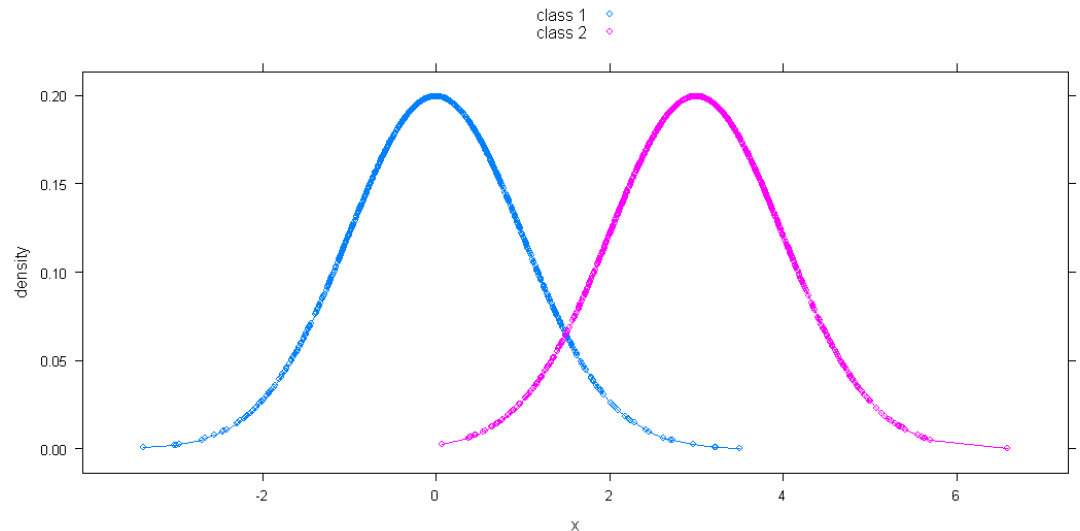
- Function mapping from pixels to object identity is complicated
- Series of hidden layers extract increasingly abstract features



Zeiler and Fergus 2014

# Generative Models

- Bayes
  - Naïve Bayes
- Graphical models
  - Belief networks



# Remember...

- No classifier is inherently better than any other: you need to make assumptions to generalize
- Two components of the error
  - Bias: due to over-simplifications
  - Variance: due to inability to perfectly estimate parameters from limited data

# Generalization



Training set (labels known)



Test set (labels unknown)

- How well does a learned model generalize from the data it was trained on to a new test set?

# Generalization

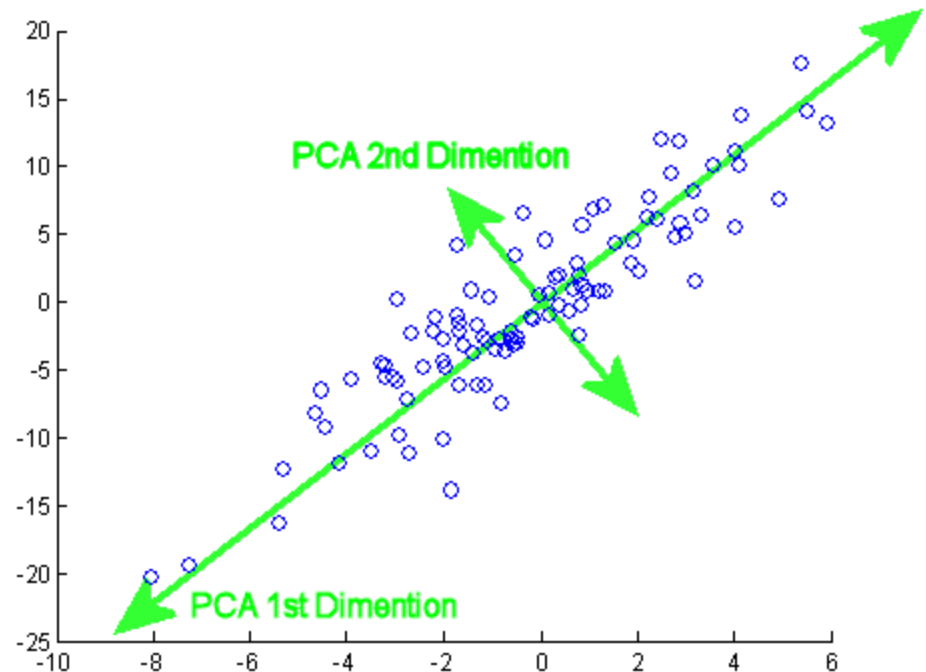
- Components of generalization error
  - **Bias:** how much the average model over all training sets differ from the true model?
    - Error due to inaccurate assumptions/simplifications made by the model
  - **Variance:** how much models estimated from different training sets differ from each other
- **Underfitting:** model is too “simple” to represent all the relevant class characteristics
  - High bias and low variance
  - High training error and high test error
- **Overfitting:** model is too “complex” and fits irrelevant characteristics (noise) in the data
  - Low bias and high variance
  - Low training error and high test error

# Clustering Strategies

- K-means
  - Iteratively re-assign points to the nearest cluster center
- Agglomerative clustering
  - Start with each point as its own cluster and iteratively merge the closest clusters
- Mean-shift clustering
  - Estimate modes of pdf
- Spectral clustering
  - Split the nodes in a graph based on assigned links with similarity weights

# Feature Extraction

- Principal Component Analysis
- Fisher Discriminant Analysis
- 





# Representing Face Images: Eigenfaces

**Q: How do we pick the set of basis faces?**

**A: We take a set of real training faces**



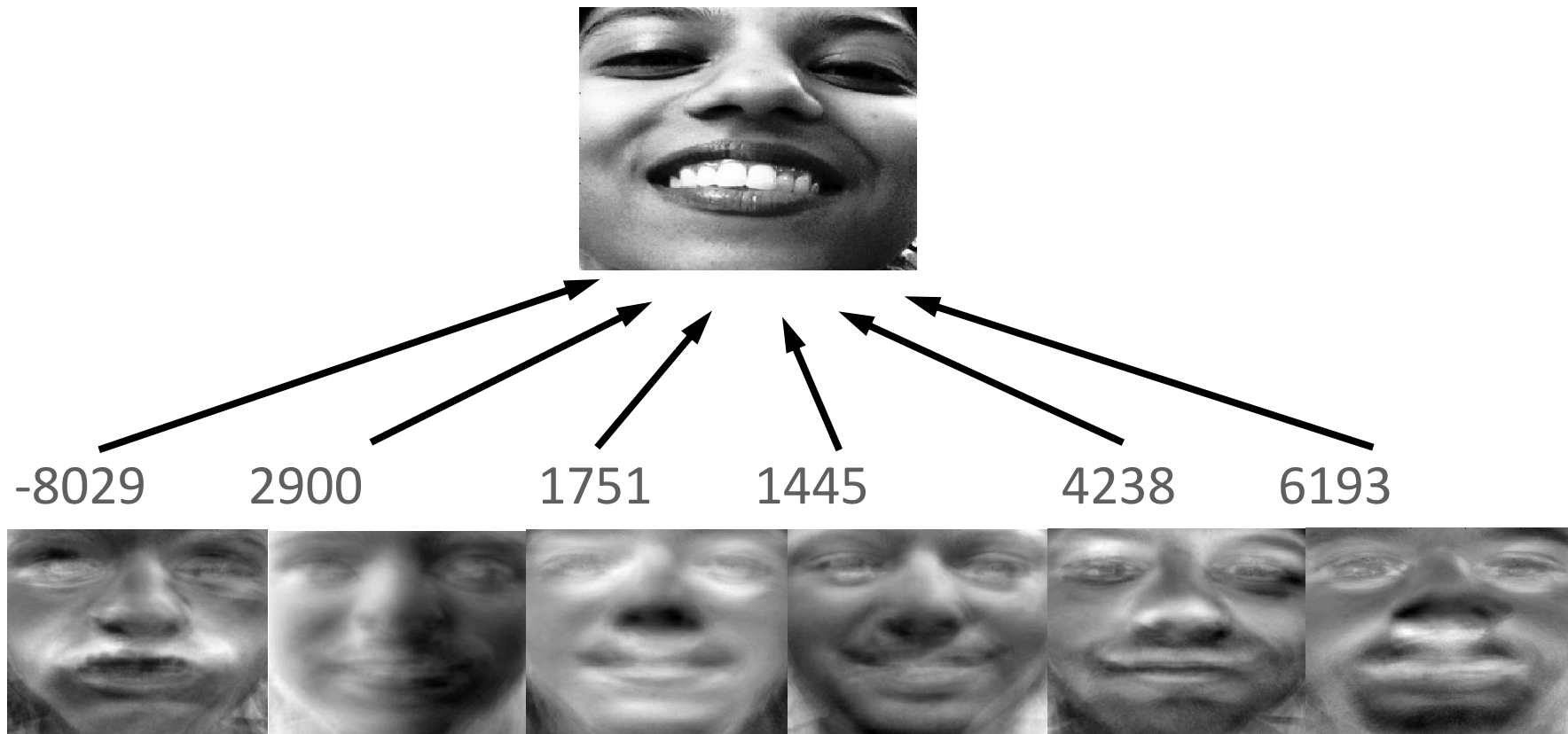
**Then we find (learn) a set of basis faces which best represent the differences between them**

**That is, apply PCA and choose top Eigen vectors (Eigen faces)**

**We can then store each face as a set of weights for those basis faces**

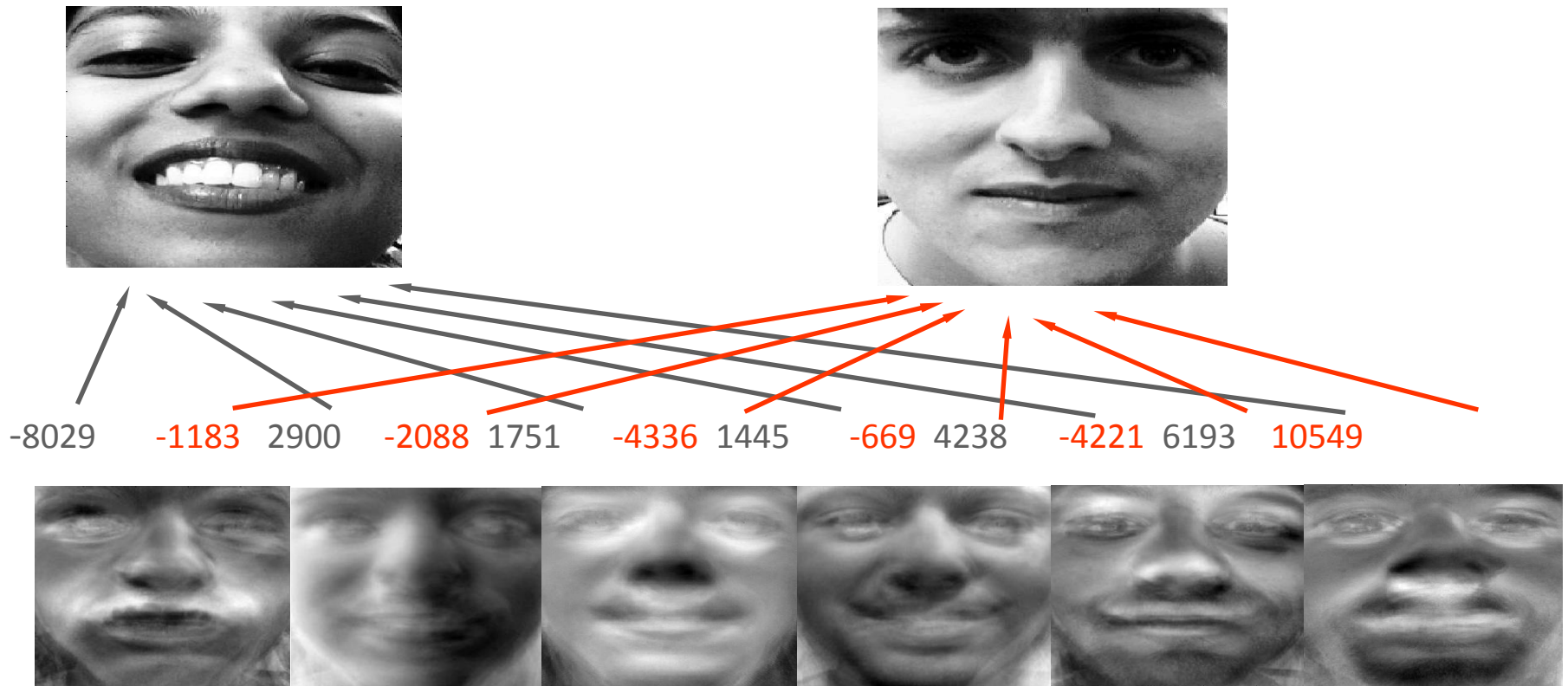
# Eigenfaces: the idea

- Think of a face as being a weighted combination of some “component” or “basis” faces
- These basis faces are called eigenfaces



# Eigenfaces: representing faces

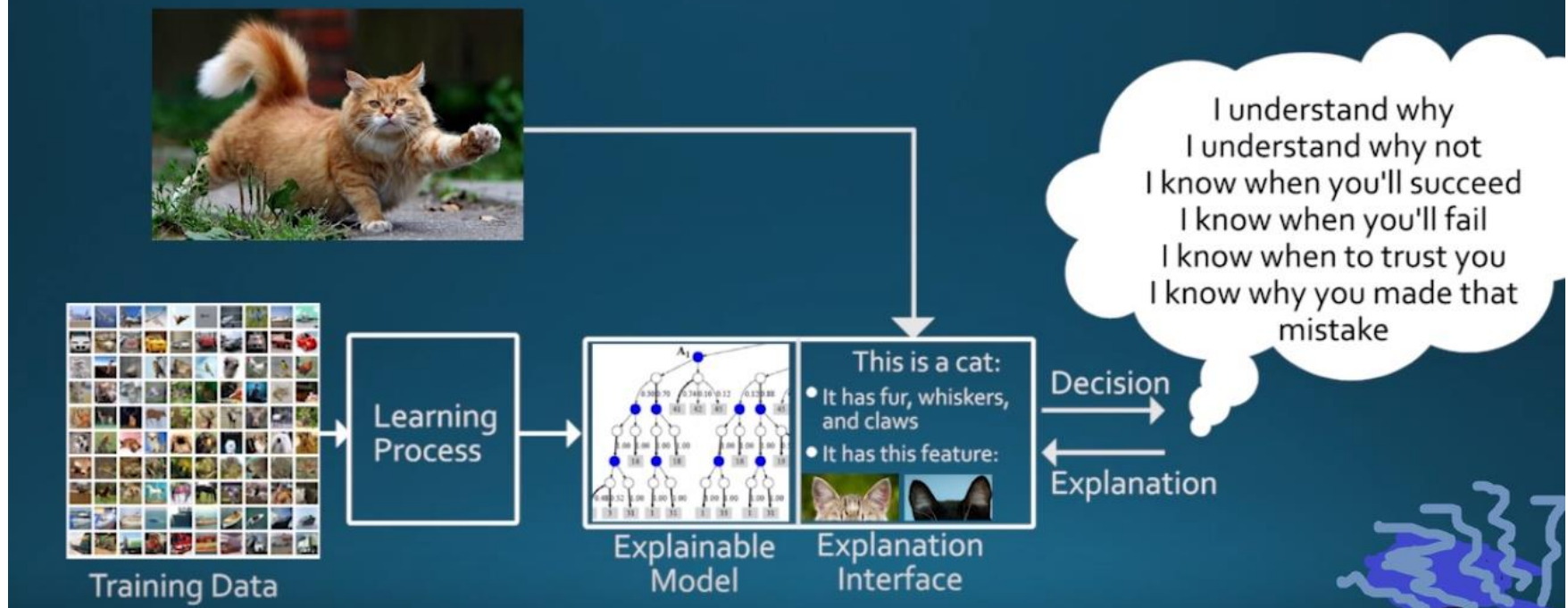
- These basis faces can be differently weighted to represent any face
- So we can use different vectors of weights to represent different faces

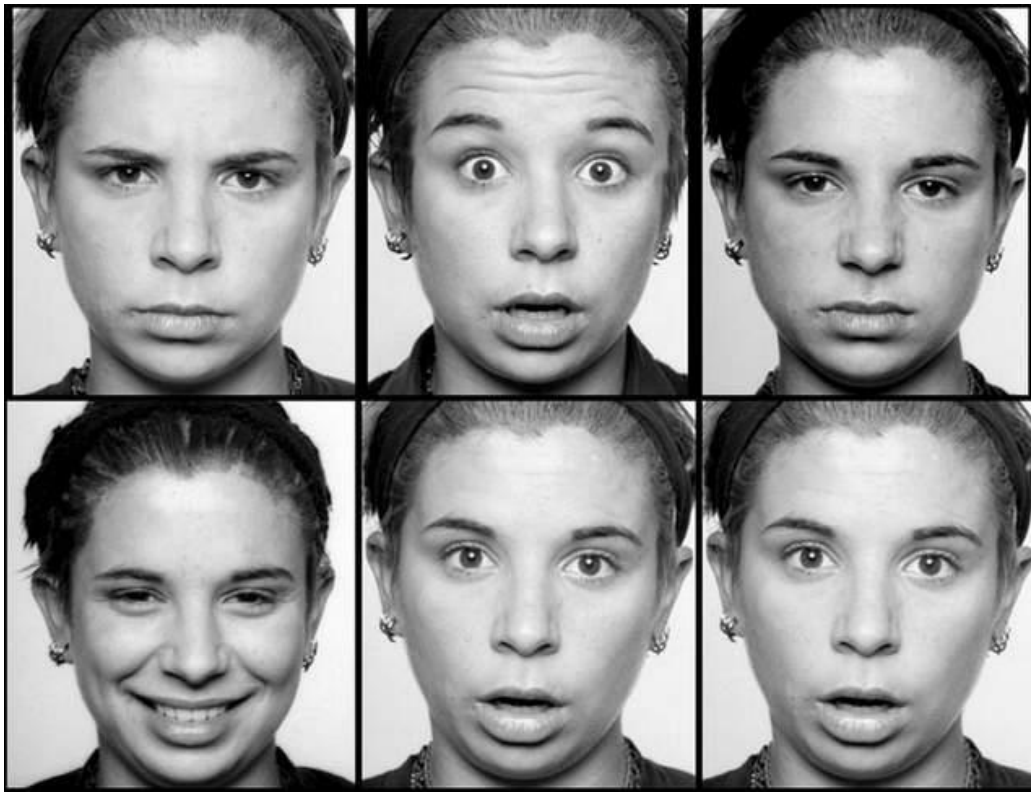


**SOME CHALLENGES ...**

- King – man + woman = Queen
- Face – emotion + surprise = Surprised face
- Interpretable models.

## Models to explain decisions

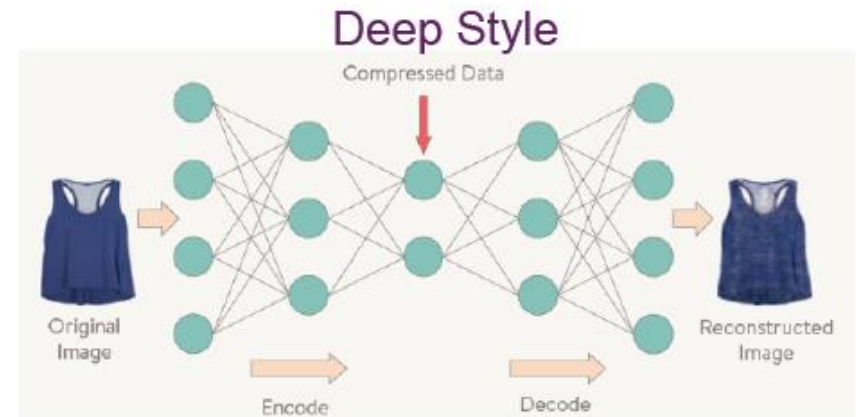






# Example of Representation Learning

- Autoencoder:
  - Quintessential example of representation learning
  - Encoder:
    - Converts input into a representation with nice properties
  - Decoder:
    - Converts the representation back to input



New designs from representation



**THANK YOU**