

# **CS4770**

## **Pattern Recognition**

### **Introduction**

**P. J. Narayanan**  
Monsoon 2005

# What is Pattern Recognition?

- Take in raw data (e.g., smell, colour, shape, frequency)
- Label or *classify* it as a **category** of patterns. (e.g., rose, ripe mango, Abdul Kalam, Hamsadhwani, etc.)
- Animals are so very good at it. This is the basis of **intelligence**.
- We go from sensory inputs (sight, sound, smell, etc.) to concepts or categories. (Giraffe, rose, etc.)

# What's in this course?

- Can computers be made to recognize or classify patterns?
- We go from data given as numbers (primarily given by a camera, microphone, thermometer) to categories given as labels or class IDs.
- We concentrate on the mathematical techniques and tools that make classifications easier in this course.

# An example

- Given data about a lot of people, we want to classify them as male and female. A simple 2-class problem.
- How do we do it?  
Use different properties or features.
- **Featue:** A relevant, defining characteristic of the object, captured using a sensor.
- Given some or all of it, how do we actually categorize it as a mango or an orange?

## Example in detail

- Use the height in centimeters to distinguish.
- **Wonder Island:** Where women are tall and men dwarves!  
<http://www.indiaglitz.com/channels/malayalam/review/7622.html>
- Find a suitable value for height below or a **decision boundary**.
- Simple: a threshold value to separate them. Can have errors.

- Likely to work in WI as the heights determine a lot. The classes are **separable** easily.
- This may not work well in other places of the world.
- Height may not be sufficient. No good decision boundary exists.
- How about using the pitch of their voice? Women will have higher pitch and men smaller.
- A threshold on this works better as males and females have distinctly different pitch profiles.

# Mango and Orange

- We know a lot about them from our experience.
- We have been **trained** by our different encounters with mangoes and oranges (and a few thousand other categories).
- An automatic system also needs to be trained by presenting it with many examples of oranges and mangoes.
- The decision rule based on the feature values will be inferred or learned by the system.

- When the same examples are presented, no error should be made.
- **Generalization:** When other mangoes and oranges are presented, no error should be made!

## Another example

- A large box, full of oranges and mangoes. Randomly pick one fruit from the box.
- Need to say if it is a mango or an orange.
- Trivial if we can take a good look at the picked fruit.
- How do we do it?  
Use colour, shape, smell, surface texture/roughness.
- **Featue:** A relevant, defining characteristic of the object, captured using a sensor.

- Given some or all of it, how do we actually categorize it as a mango or an orange?

## Example in detail

- Colour: variation for mango and orange. Shape: similar.
- Find a decision boundary.
- Simple: a line separating them. Can have errors.
- Complex: actual boundary in the space.
- Or some middle ground.

# Mango and Orange

- We know a lot about them from our experience.
- We have been **trained** by our different encounters with mangoes and oranges (and a few thousand other categories).
- An automatic system also needs to be trained by presenting it with many examples of oranges and mangoes.
- The decision rule based on the feature values will be inferred or learned by the system.

- When the same examples are presented, no error should be made.
- **Generalization:** When other mangoes and oranges are presented, no error should be made!

# Pattern Recognition System

- A mechanical system has several steps: sensing, segmentation or grouping, feature extraction, classification, evaluation.
- Sensing: Collecting the raw information about it. Camera, Smell sensor, microphone, etc., could be used.
- Segmentation: Separate the possible fruits from non-fruits! Hard.
- Feature extraction: Smell, colour etc. are represented as

numbers, which may need to be computed from the data given by the sensors.

- Classification or Decision making: Assign a label to the object based on the feature values.
- Evaluation: Look at the error. Use other context.

# Design of a PR System

- Decide on the sensors for data collection.
- Decide on the features to use. They should be relevant for the task. They should be easy to compute from the sensors. Use all prior or expert knowledge for this.
- Select a model. Decide on what would be the structure (mathematical or otherwise) underlying the objects we are trying to classify.
- Training: Get enough inputs whose categories are known and initialize the model parameters.

- Evaluation: How does the system do on the training examples? How does it do on other known examples? Is the model adequate? Are the features good?

# Learning

- All training involves some learning of the model parameters.
- **Supervised Learning:** A teacher provides a category label along with the data for each pattern.
- **Clustering/Unsupervised Learning:** No teacher to present the correct category label. The “natural” differences between the categories are discovered from the data itself.

- **Reinforcement Learning:** Category label is not presented with each pattern. However, a critic will tell the system if the classification is right or wrong. This information can be used to improve the model

# Topics

- Statistical Pattern Recognition: Underlying statistical model is known.
- Parameter Estimation: Computing the statistical model from examples.
- Linear Classifiers: Linear functions to discriminate between different classes.
- Neural networks: A generalization of the above.
- Non-metric models: Clustering and Unsupervised Learning

# Course Plan

- 5-6 Homework assignments: 10-15%
- 2 Tests: 30-40%
- Project: 20-30%
- Final exam: 25-35%
- Others: 5-10%

Textbook: **Pattern Classification**. Duda, Hart, and Stork. 2001. John Wiley. Indian Edition available.