Introduction

- The purpose of the project is to enable the student to get hands-on experience in the design, implementation and evaluation of pattern recognition algorithms.

Minimum Requirements

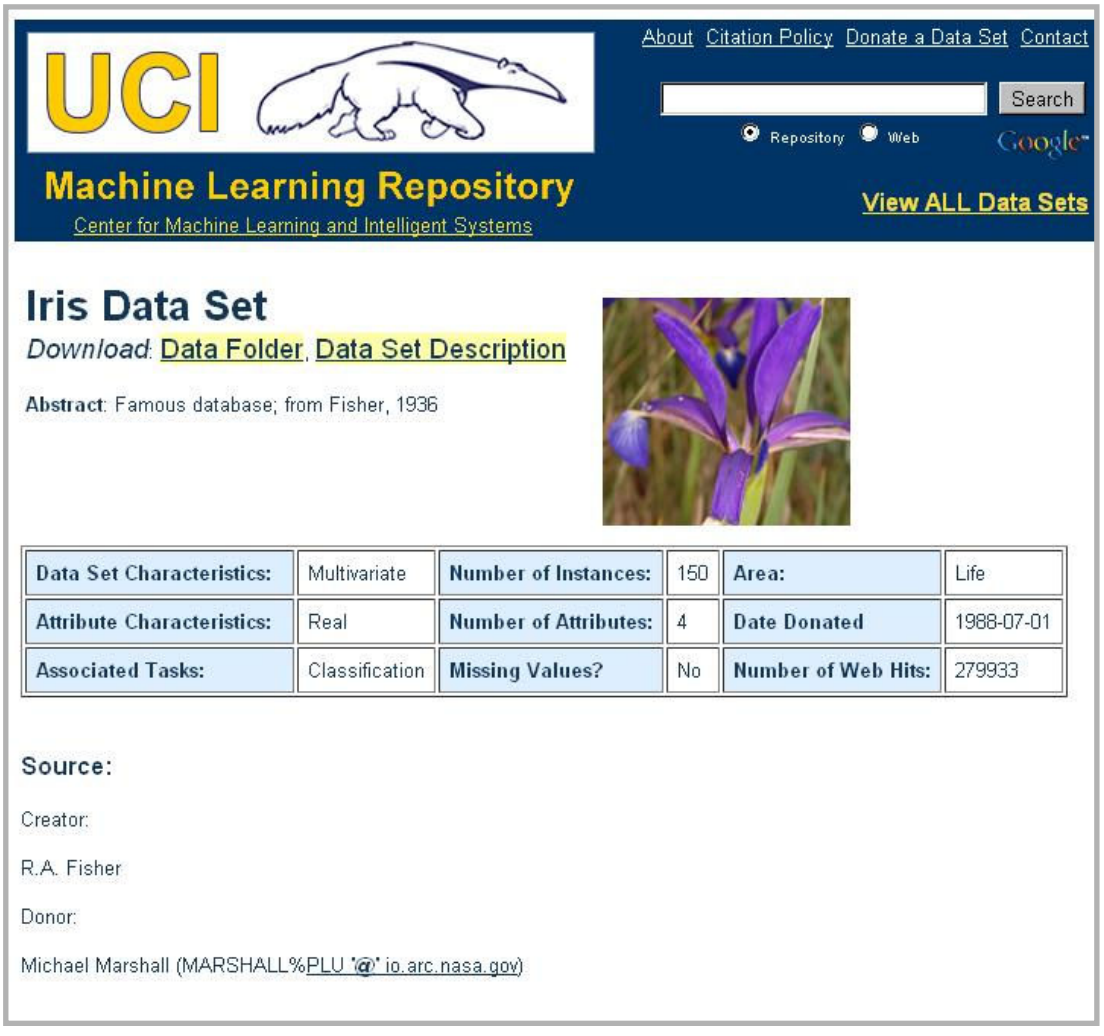
- In this project, you have to…
  - Find out the best one out of 5 classifiers to solve one selected problem
    - Evaluation their performances
  - Additional task (Optional)
    - Feature Selection
    - Sample Selection
    - New learning method
    - Anything you think it can help to satisfy the users

Topics you can select

- Spambase
- Blood Transfusion Service Center
- Breast Cancer Wisconsin (Diagnostic) Data Set
- Yeast Data Set
- Glass Identification Data Set
- **The ORL Database of Faces (difficult)**
- **PIE Database (very difficult)**

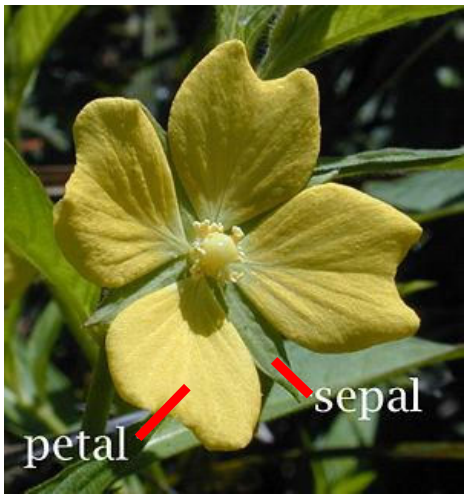
- If you want to do other interesting topic (assume that you can find the related dataset), please let me know

Iris



Iris Example

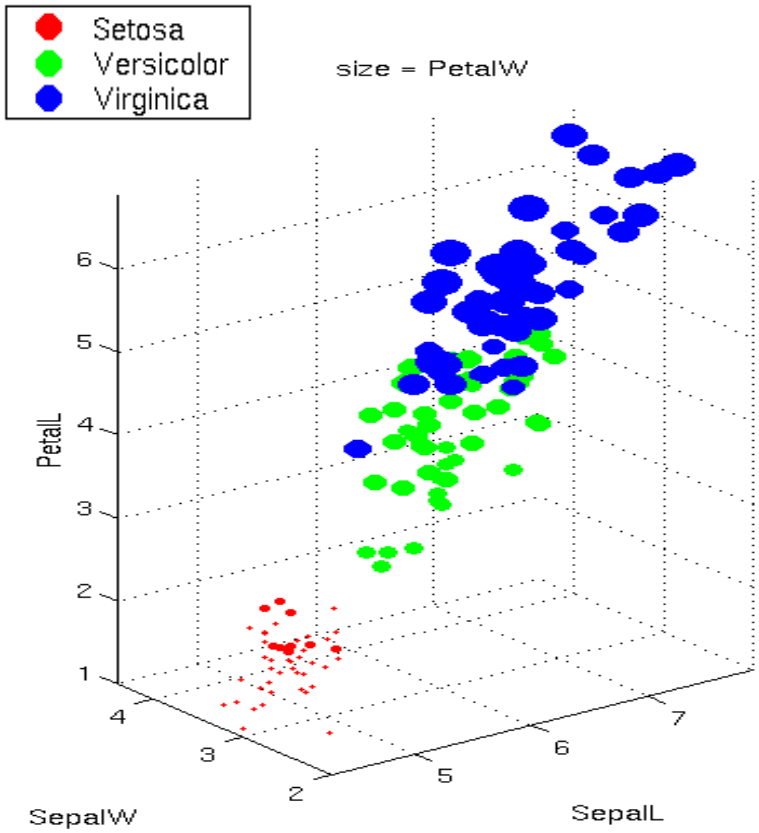
- **Iris Data Set**
  - 4 Features
    - sepal length in cm
    - sepal width in cm
    - petal length in cm
    - petal width in cm
  - 3 Classes
    - Iris Setosa
    - Iris Versicolour
    - Iris Virginica



Iris Example

```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5.4,3.7,1.5,0.2,Iris-setosa
4.8,3.4,1.6,0.2,Iris-setosa
4.8,3.0,1.4,0.1,Iris-setosa
4.3,3.0,1.1,0.1,Iris-setosa
5.8,4.0,1.2,0.2,Iris-setosa
5.7,4.4,1.5,0.4,Iris-setosa
```

.
.
.



Requirements

- Preprocessing:
  - All features are necessary and useful?
    - More is better?
  - All samples are necessary and useful?
    - More is better?
  - If some features not a number, what should we do?
    - E.g. “Female”, “Male”
    - “Good” “OK” “Bad”
  - Normalization is needed?
    - Feature A is from -100 to 100
    - Feature B is from 0.5 to 2.7
  - More???

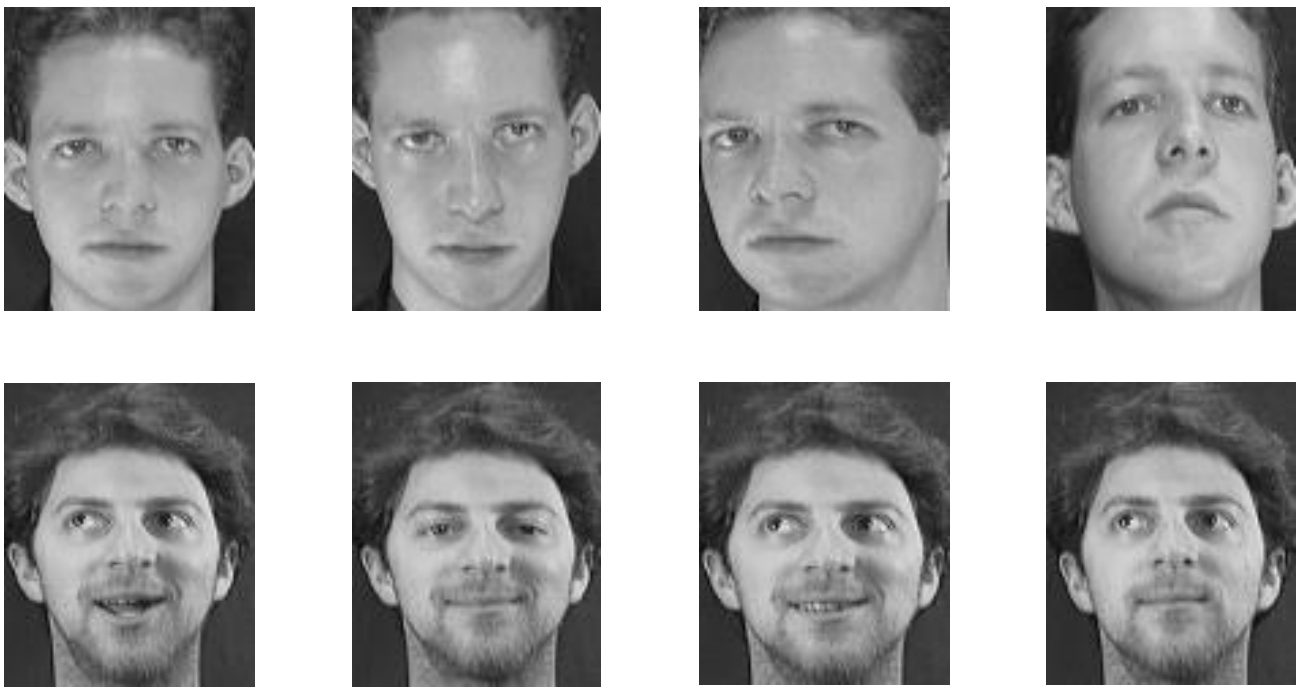
Requirements

- Classifier
  - Classifier Selection
    - Which kind(s) of classifier is the most suitable for the question? Why?
  - How to choose the parameter(s)?
    - E.g. Number of Layer in MLP?
  - How to train the classifier?
    - Time complexity…
  - If the trained classifier is not good, what should we do?
  - Improve the learning algorithm?
  - More???

Requirements

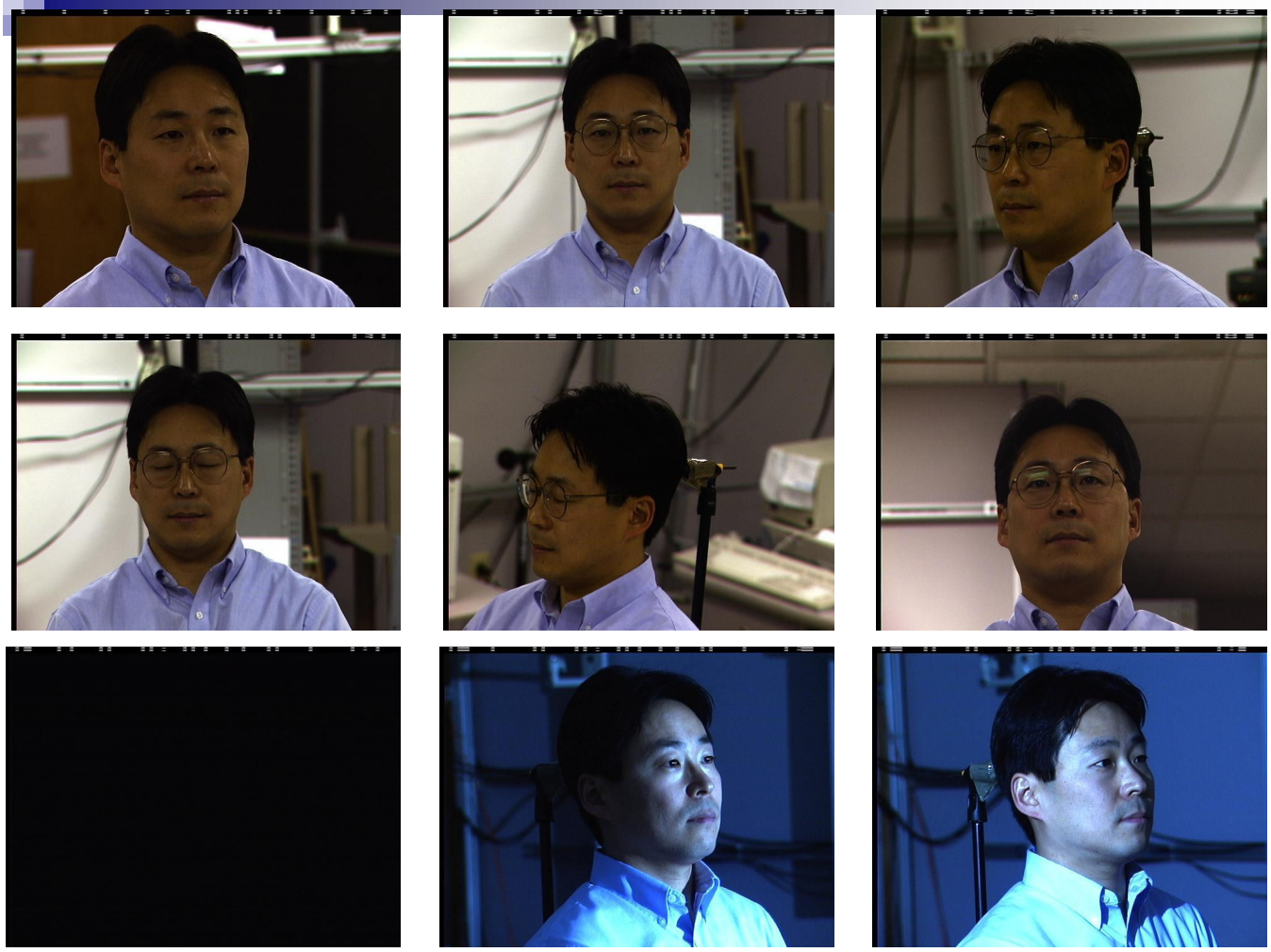
- Analysis
  - Which classifier(s) is better?
    - How to compare them?
  - Are your classifiers are good enough?
    - Anyway can do better if you have more resources?
  - Any useful information can be found rather than telling users you trained a 99.99% accurate classifier?
    - Help user to understand more the problem
  - Similarity special situations? E.g. Noise
  - More???

ORL Database of Faces (Difficult)



PIE Database (Very Difficult)

- Lighting
- Angle
- Expression



Group

- Each group should have **FIVE** students
- One of group member is leader
- **Send an email** to **Mr. Zhimin He** (531981194@qq.com) with the following information **by 6-April-2011**
  - Student ID
  - Class
  - Name
  - Mobile (only for group leader)
- Email Title: **"PR Course: Project Group"**

Submission

- **Report (in English)**
  - Explain what you have done
  - Record every steps and thinking
- **Give a presentation (in English)**
  - 10 minutes
- The presentation date =
  the submission due date =
  9-June-2011