

Practice Class 1

Introduction for Machine Learning

2013.03.26

Eun-Sol Kim



Contents

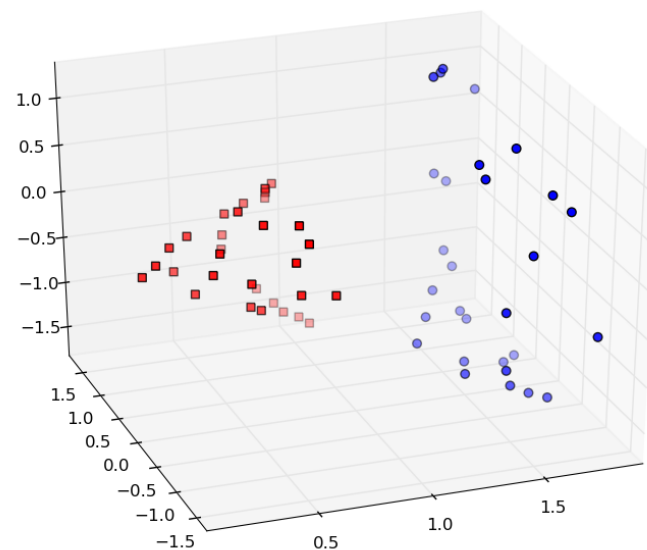
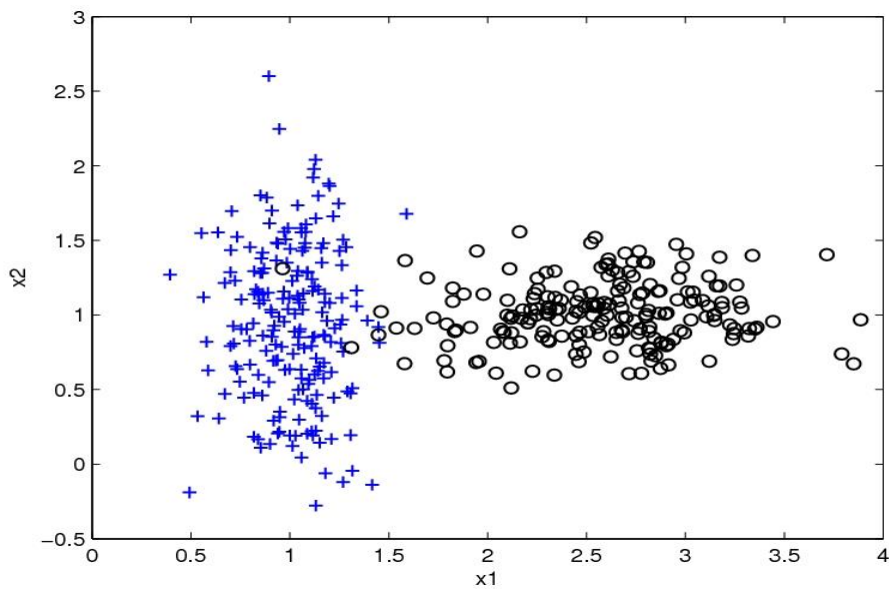
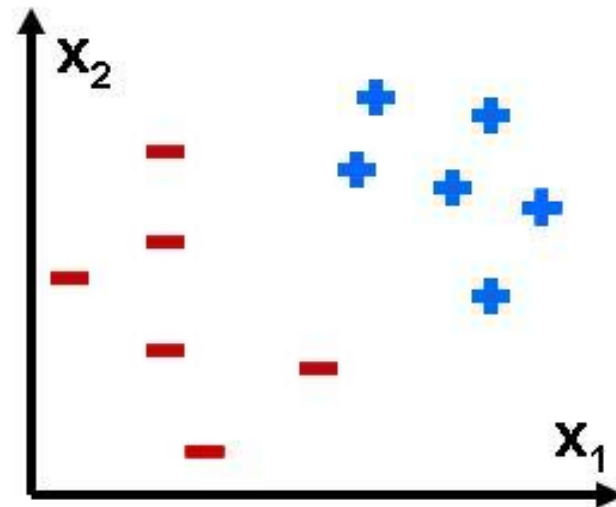
- Goal for practice class
- About the poster presentation

- What is the classification?
- Data, Feature, Attribute, Class label, Algorithms
- Examples for open dataset

Goal for the practice classes

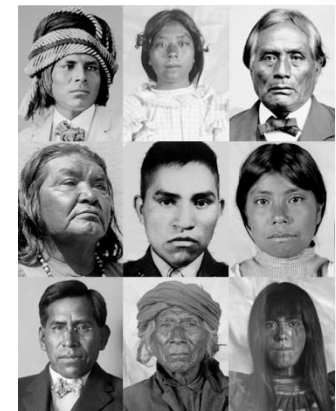
- To have experience about all steps of 'Machine Learning'
 - Define a problem
 - Make or Fine dataset
 - Experiment
 - Analysis

Classification



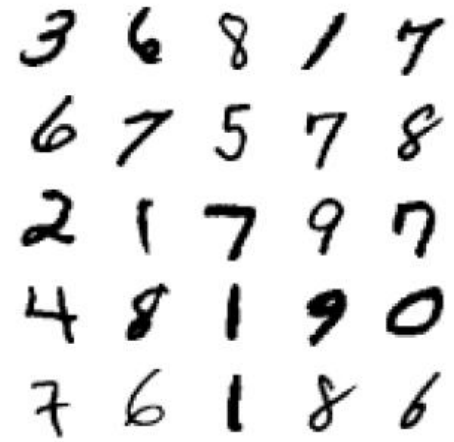
Example Dataset #1: Pima Indians Diabetes

- Description
 - Pima Indians have the highest prevalence of **diabetes** in the world
 - We will build **classification models** that diagnose if the patient shows signs of diabetes
 - <http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>
- Configuration of the data set
 - 768 instances
 - 8 attributes
 - age, number of times pregnant, results of medical tests/analysis
 - all numeric (integer or real-valued)
 - Also, a discretized set will be provided
 - Class value = 1 (Positive example)
 - Interpreted as "tested positive for diabetes"
 - 500 instances
 - Class value = 0 (Negative example)
 - 268 instances



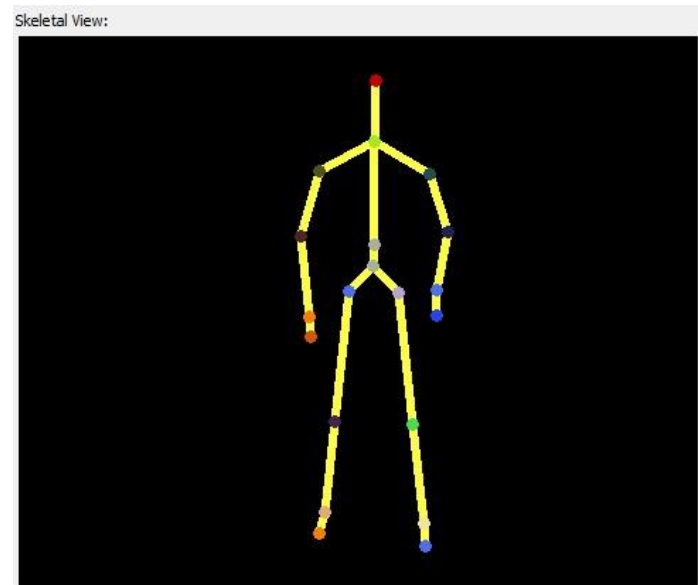
Example Dataset #2: Handwritten Digits (MNIST)

- Description
 - The MNIST database of handwritten digits contains digits written by office workers and students
 - We will build a **recognition** model based on classifiers with the reduced set of MNIST
 - <http://yann.lecun.com/exdb/mnist/>
- Configuration of the data set
 - Attributes
 - pixel values in gray level in a 28x28 image
 - 784 attributes (all 0~255 integer)
 - Full MNIST set
 - Training set: 60,000 examples
 - Test set: 10,000 examples
 - For our practice, a reduced set with 800 examples is used
 - Class value: 0~9, which represent digits from 0 to 9



Dataset you can use

- Default dataset in Weka
- UCI repository
 - <http://archive.ics.uci.edu/ml/>
- CIFAR dataset
 - <http://www.cs.toronto.edu/~kriz/cifar.html>
- Make your own dataset
 - For example, motion data



Experiment with Weka

