

Introduction

Dr. Fayyaz ul Amir Afsar Minhas

PIEAS Biomedical Informatics Research Lab Department of Computer and Information Sciences Pakistan Institute of Engineering & Applied Sciences PO Nilore, Islamabad, Pakistan <u>http://faculty.pieas.edu.pk/fayyaz/</u>





Apples

Oranges

What is this?



Paintings by two different painters







How many categories (clusters) are there?



Find the odd one out!





Predict the series

• 1,1,2,3,5,8,13,...



Question?

- Consider the vectors
 - $-X_1 = [1 \ 2 \ 1 \ 4]^T$
 - $-X_2 = [2 4 2 4]^T$
 - $-X_3 = [0 \ 0 \ 0 \ 4]^T$
 - $-X_4 = [3 6 3 4]^T$
 - $-X_5 = [4 8 4 4]^T$
- To store each vector, how many dimensions (or variables) do we need?



Questions

- How were you able to recognize that the object shown was indeed an apple?
- How were you able to discriminate between the paintings from two different painters?
- How were you able to find out the different types of apples in the picture?
- How did you manage to find the next number in the series?
- How were you able to find which dimension was redundant?
- How were you able to find the odd one out?
- Learning to drive / write?

Classification Classification

Clustering

Regression

Dimensionality Reduction

Anomaly Detection

Reinforcement learning

What is Machine Learning?

- Computers are _____
 - Dumb
- Making a machine (computer) perform the same tasks which you have just done is called

- Artificial Intelligence

- If you learn to do these tasks using existing data, then this is called _____
 - Machine Learning

Well Humans learn, don't they?

- Of course!
- In many many ways!!
 - Instruction based learning
 - Rote Learning
 - Informal Learning
 - Active Learning
 - Enculturation
 - Experience based learning





Example: Human Learning

- Science is based on developing and testing hypothesis that "explain" our universe
- For example:
 - Newton's Formula F = ma explains the motion of an object of mass m when a force F is applied to it
 - Scientists observed that Newtonian mechanics does not "explain" the motion of mercury properly
 - This led to the development of theory of relativity by Einstein which explains it!!
- We constantly try to develop and refine models of the world and the universe
- However <u>sometimes</u> it gets hard!



Why do we need computers?

- ...ATTCGAGGATTACACCGTAAGAAATTT...
- ...ATCGCCTGATTACATATATACCGTTGG...
-AGATTAAATCGTTCGATTCACATTGAC
- Deduction vs. Induction Reasoning
- High dimensions
- Required Reading
 - Halevy, Alon, Peter Norvig, and Fernando Pereira.
 "The Unreasonable Effectiveness of Data." *IEEE Intelligent Systems*, 2009.
 - Quiz in next class

Machine Learning

 A unique junction of computer science, applied mathematics statistics and the world!



Machine Learning Overview

- Types
 - Supervised
 - Classification, Regression, Reinforcement learning
 - Unsupervised
 - Classification, Visualization, Representation
 - Semi-Supervised
- Approaches
 - Discriminative
 - Generative
- Classes of Algorithms
 - Distance Based
 - Neural Networks
 - Deep Learning
 - Large Margin Methods (Kernels)
 - Ensemble Techniques
 - Logic Based
 - Probabilistic, Bayesian Networks

Philosophies

- Applied
 - Pattern Recognition
- Theoretical
- Developmental
 - Making new learning algorithms
- Hybrid

Applications...

- Speech recognition
- Handwritten character recognition
- Face detection and tracking
- Face recognition Medical research
- Machine vision
- Target recognition
- Radar image analysis
- Geographical information processing
- Image segmentation
- Credit scoring Nuclear reactors
- Astronomy
- Financial engineering
- Financial time series prediction
- Drug design
- Cancer diagnosis
- Text analysis

This course

• Aims

- Concepts of Machine Learning
- Hands on experience
- Experiment design strategies for machine learning
- Contents
 - Weeks 1-5 (Assignments 1 and 2)
 - Introduction: Machine Learning Framework, Python
 - Classification: kNN, Neural Networks, Logistic Regression, SVM (Linear and Kernels)
 - Weeks 6 (Projects Begin)
 - Experiment Design: Evaluation Metrics, Cross-Validation, Model Selection
 - Week 7-8 (Research Assignment)
 - Theoretical Foundations: Naïve Bayes, Capacity, VC Dimensions, Structural Risk
 - Week 9-10 (Assignment: PCA & LDA)
 - **Dimensionality Reduction**: Linear (PCA, LDA, Kernel PCA)
 - Week 11-12 (Assignment: Regression)
 - **Regression**: LS, PLS, SVR
 - Week 13 (Assignment: Clustering)
 - **Clustering**: Hierarchical, k-means, SVC, Biclustering
 - Week 14-16 (Tentative Midterm and Project Presentations Due)
 - Feature Selection and Mining / Reinforcement Learning
 - Advanced Topics
 - **Application**: Projects

Evaluation

- Quizzes/Reading: 15%
- Assignments: 20% (tentatively 5 Assignments)
 - 1 review report and presentation
 - All other programming based
 - Should not take more than 4 hours of effort in each
- Project: 10%
 - Weekly progress and presentation, In groups of 2
- Mid Term: 5% (to prep you for the final!)
- Bonus Marks: Class Questions, Use of JULIA
 - <u>http://julia.readthedocs.org/en/release-0.4/</u>
- Final Exam: 50%
 - Including at least 10% of questions related to projects
- Effort Required: Avg. 2 hours per 1 class hour

Assignments

- Making a kNN classifier
 Experiments in dimensionality
- Implementation of a simple linear classifier
 - Using Linear Discriminant Functions
 - SVM Optimization
- Coding and Applying PCA and LDA
- Clustering / Regression
- Research Assignment
 - Ensemble Methods
 - Regression Comparison ...

Logistics

- Course Webpage
 - Piazza
 - <u>https://piazza.com/pieas.edu.pk/other/cis621/home</u>
 - Please register, signup or send me an email so
 - afsar at pieas dot edu dot pk
 - Please use a single email address for all interaction and be sure to check it daily for updates
- Office Hours
 - (Free Tuition Time!) Tuesdays 1140-1230 in B-216
 - By Email Appointment
- Attendance: PIEAS Policy

Resources

- Books
 - \\172.30.10.2\FacultyShare\Fayyaz ul Amir Afsar
 Dr\CIS621 Machine Learning
- Python Help
 - Faculty Share
 - \\172.30.10.2\FacultyShare\Fayyaz ul Amir Afsar Dr
 - \\172.30.10.2\FacultyShare\Fayyaz ul Amir Afsar Dr\PYTHON
- Online Help
 - Scikit: <u>http://scikit-learn.org/stable/tutorial/basic/tutorial.html</u>
 - Scipy: <u>http://www.scipy-lectures.org/</u>

Self-Learning Requirements

- Python
 - Install Anaconda Python Distribution in Windows or Spyder in Ubuntu Linux
 - What you need to understand:
 - Installation and Administration
 - Using package manager (pip, easy_install, conda)
 - Basic Constructs: Variables, Control Flow, Object Oriented concepts, Mutable and Immutable Types, Lists and Dictionaries
 - Using Scipy (Matplotlib for plotting and Numpy)
 - Debugging (pdb)
 - Primary package: scikit-learn
- Reading pointers will be given
 - Can discuss issues in the office hours

To Do

- Reading
 - Required
 - Halevy, Alon, Peter Norvig, and Fernando Pereira. "The Unreasonable Effectiveness of Data." *IEEE Intelligent Systems*, 2009.
 - "Machine Learning." Wikipedia, the Free Encyclopedia, January 20, 2016.

https://en.wikipedia.org/w/index.php?title=Machine_learning&oldid= 700785899.

- Quiz Next Lecture
- Optional
 - Domingos, Pedro. "A Few Useful Things to Know About Machine Learning." Commun. ACM 55, no. 10 (October 2012): 78–87. doi:10.1145/2347736.2347755.
- Python
 - Installation
 - Begin reading Lectures (1.X and 2.X)

End of Lecture-1

We want to make a machine that will be proud of us.

- Danny Hillis

CIS 621: Machine Learning