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
http://faculty.pieas.edu.pk/fayyaz/

This course

- **Aims**
  - Concepts of Machine Learning
  - Application of machine learning in Biomedical informatics
  - Experiment design strategies for machine learning
  - Development of required research techniques for a career in MLIB
  - Research Output: Projects / Papers
- **Contents**
  - **Week 1: Introduction:** Machine Learning Framework, Python/Julia
    - A few useful things to know about machine learning
    - Machine learning in Bioinformatics: Challenges
    - Application areas
      - » CRISPR: Acr, Off-target effects, Guide Design
      - » Protein Interaction Prediction: Interface prediction, Binding Affinity Prediction
      - » Anti-microbial peptide prediction and design
      - » Hemometer, correlative machine learning applications
      - » Heterogenous Machine Learning: Different features for different examples
      - » Data Minimal AI
      - » A review of features used in proteins
      - » A review of how we should do machine learning in bioinformatics
  - **Weeks 2-3: Fundamentals of Machine Learning: Classification**
    - Discriminant Based Classification, Perceptron , SVM
    - Kernel Methods in Biomedical Informatics
    - Assignment-2: (A) Implement perceptron, (B) SVM, (C) Kernels for Different Problems
  - **Week 4: Experiment Design in ML for Bioinformatics**
  - **Week 5: Fundamentals of Machine Learning: Theoretical Foundations (SRM)**
  - **Week 8-12: Advanced Learning Machines**
    - **Regression [8], Ranking[9], Multiple Instance Learning [10]**
    - **Introduction to:** Online Learning, Transductive Learning, Transfer Learning, Multi-task Learning, Mult-label Learning, Recommender Systems, Self-Taught Learning, High Dimensional Feature Mappings [10]
    - **Structured Output Learning [11]**
    - **Correlation Filters [12]**
  - **Week 13: Weak Supervision Techniques**
    - Learning using privileged information
  - **Week 14: Probabilistic Soft Logic**
  - **Week 15-16: Deep Learning in Biomedical Informatics**
    - **Introduction to existing methods**
    - **Deep learning techniques for protein modeling**

Books/Refs

- Kernel Methods for Pattern Analysis
- Understanding Machine Learning: From Theory to Algorithms by Shai Shalev-Shwartz and Shai Ben-David (excellent intro to structured prediction)
- Learning with Kernels
- SEQUENCE-BASED PREDICTION OF RESIDUE-LEVEL PROPERTIES IN PROTEINS in MACHINE LEARNING IN BIOINFORMATICS
- Neural Networks: Tricks of the Trade http://ciml.info/
- Deep Learning

Evaluation

- **Assignments: 35%**
  - **Critique Writing**: Review a paper, write a summary of it and identify flaws and potential improvements in it [5 in total 20 Marks]
  - **Programming based** (tentatively 3 Assignments) [15 Marks]
    - Should not take more than 4 hours of effort in each
- **Project (individual: 35%):** You can take your FYP if it is related and your supervisor agrees
- **Mid Term/Quiz**:5% (to prep you for the final!)
- **Final Exam:** 50%
  - Written Exam: 30%
  - Computer Exam: 20% (project work evaluated)
- Presentation in Group Meeting / 3rd Class
- No Extensions in deadlines
- Bonus Marks: Class Questions, Use of JULIA
  - http://julia.readthedocs.org/en/release-0.4/
- Effort Required: *Avg.* 2 hours per 1 class hour

Assignments

- Assignment-1
  - Making a kNN classifier (Experiments in dimensionality)
  - Implementation of a simple linear classifier
  - Theoretical Questions
- Assignment-2
  - Implement Kernel Methods
- Assignment-3
  - Clustering / Regression
- **Assignment-4 (Begins from Day-1)**
  - **Write a review/critique of literature for your project**
  - **To be submitted with your sessional result**
  - **Evaluation by Presentation**

Tentative Critique Papers

- [ALL] Do we need hundreds of classifiers to solve real world classification problems?
  - Are Random Forests Truly the Best Classifiers?
- [ALL] Design and Analysis of Classifier Learning Experiments in Bioinformatics: Survey and Case Studies
  - A systematic Comparison of Supervised Classifiers
- [ALL] Protein-protein binding affinity prediction from amino acid sequence
  - Comment on 'protein–protein binding affinity prediction from amino acid sequence'
  - Response to the comment on 'protein-protein binding affinity prediction from amino acid sequence'
  - PRODIGY: a web server for predicting the binding affinity of protein–protein complexes
  - Computational Approaches for Predicting Binding Partners, Interface Residues, and Binding Affinity of Protein–Protein Complexes
- [ALL] Deep Learning (http://www.nature.com/nature/journal/v521/n7553/full/nature14539.html)
  - Deep Learning in Bioinformatics (https://arxiv.org/pdf/1603.06430)
  - Deep learning for computational biology
  - The path from big data to precision medicine
  - Use of big data in drug development for precision medicine
  - Predicting the errorImproving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks
  - EP-DNN: A Deep Neural Network-Based s of predicted local backbone angles and non-local solvent-accessibilities of proteins by deep neural networks
  - Global Enhancer Prediction Algorithm
- 1 Written Page For Each Critique With References
- 10+5 Minute Presentation
  - Submission 1 day before the evaluation

Logistics

- Course Webpage
  - Piazza
  - 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 1030-1130 in B-216
- 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: http://scikit-learn.org/stable/tutorial/basic/tutorial.html
  - Scipy: http://www.scipy-lectures.org/
  - Julia: http://julia.readthedocs.org/en/release-0.4/

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
    - 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.
    - “Machine Learning.” *Wikipedia, the Free Encyclopedia*, January 20, 2016. https://en.wikipedia.org/w/index.php?title=Machine_learning&oldid=700785899.
  - Optional
    - Halevy, Alon, Peter Norvig, and Fernando Pereira. “The Unreasonable Effectiveness of Data.” *IEEE Intelligent Systems*, 2009.

Critiquing

- Identify the main question
  - Why is this question being asked
  - What is the motivation
- What is the most critical aspect of the design of the experiment
  - Identify the comparison
  - Identify the control conditions
  - Potential Shortcomings in experiment design
- Classic problems in analysis/inferences/conclusions
  - Identify the primary conclusion
  - Does the data really point to that conclusion?

Critiquing Specific to Machine Learning

- Practicality
- Data Collection
- Performance Evaluation
- Application of the machine learning model
- Comparisons

Grading of critiques

- Summary [4]
  - Effectively summarize the paper in a maximum of two-three sentences [1]
  - Provide context of the proposed method [1]
  - Finding papers relevant to the given paper [1]
  - Pose an ORIGINAL question or problem related to the overall goal of the paper [1]
- Proposed Methodology [4]
  - Identification of a valid flaw in the design or proof of the proposed method OR Demonstrate that the design or proof of the proposed method is convincing [1]
  - Suggest reasonable improvement to the proposed method [1]
  - Propose an original and plausible application [1]
  - Meaningful comparison of the proposed method to that of another paper [1]
- Experiment Design [2]
  - Identify a valid shortcoming about experimental design OR demonstrate that the experimental design is convincing [1]
  - Suggest a reasonable improvement in experiment design [1]
- Results [2]
  - Identify a valid shortcoming in results Or explain how the results are convincing [1]
  - Meaningful comparison of the state of the art results [1]
- Presentation [3]
  - Grammar, writing skill and use of scholarly tone [2]
  - Presentation [1]

- Total Marks: 15

How to Critique

- Marshall, Gill. “Critiquing a Research Article.” *Radiography* 11, no. 1 (February 2005): 55–59. doi:10.1016/j.radi.2004.09.001.

End of Lecture-1

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

- Danny Hillis