FOUNDATIONS OF MACHINE LEARNING (AI60203)

Announcements

Course Logistics

Course Logistics

Teachers: Mahesh Mohan M R
Ph.D Teaching Assistants: Chhavi Chaudhary and Subhankar Maity
References
• Introduction to Machine Learning, Ethem Alpaydin, MIT Press (available online)
• Machine Learning, Tom M Mitchell, Mc Graw Hill Education (available online)
Theory
• Monday – 11:00-11:55 AM, Tuesday – 8:00-9:55 AM
• Evaluation: Class Test-1, Midsem, Class Test-2, and Endsem (10%+30%+10%+40%)
Programming
• Four sessions on Saturdays (will be announced)
Continuous Evaluation over four sessions (10%)
Office Hours: Fridays 4-5 PM based on appointments
Plagiarism: No tolerance policy. Binary marking (both parties).
Attendance: Compulsory to avoid deregistration.

Course Logistics

Introduction to Machine Learning

Intro to ML

Motivations of ML. Why ML is trending now?
Classification of ML: Supervised, Unsupervised and Reinforcement Learning
Algorithm and its ineffectiveness for problems intuitive to humans (e.g., digit classification)
Basics of Supervised Learning: Labelled Data, Model, Loss, and Parameter Optimization
Basics of Unsupervised Learning: Clustering and Association
Basics of Reinforcement Learning

Intro_to_ML

Reading Exercise: Intro to ML by Ethem Alpaydin (first reference): Sec 1.1, Secs. 1.2.1 - 1.2.3.

Reading Exercise: Intro to ML by Ethem Alpaydin (first reference): Secs. 1.2.4 - 1.2.5.

K Nearest Neighbor

Instance based vs Model based Learning
Local vs Global Approximation of Target

K Nearest Neighbor for Classification and Regression
Effect of K (less K -- > sensitive to noise and large K --> sensitive to possibly irrelevant inputs)
Weighted KNN for Classification and Regression
Issues of KNN
Measurement Scales (Solution: min-max or z-score normalization)
Curse of Dimensionality
Expensive and Storage Need

knn_full

Reading Exercise: Machine Learning, Tom M Mitchell (second reference): Secs. 8.1-8.2.

Linear and Logistic Regression

Motivation of Linear Models
Prerequisites of Linear Algebra
Vector Inner Product (for linear models with a single scalar output)
Matrix-vector multiplication (for linear models with multidimensional outputs)
Inner Product perspective --> for a single op value
Outer product perspective --> for the entire op
Perceptron
History
Pre-inner product Interpretation
Post-inner product Interpretation
Optimization
Gradient descent
Batch Gradient Descent
Stochastic Gradient Descent
Minibatch Gradient Descent
Closed Form Solution
Prerequisites of Probability
Probability and Random Variables
Probability Distribution (Joint, Marginal, and Conditional)
Bayes' Theorem
Independence
Mean, Variance, and Covariance
Logistic Regression
Linear vs Logistic Regression

Linear_algebra_basics

Reading Exercise: Linear_algebra_basics

Perceptron

Reading Exercise: Machine Learning, Tom M Mitchell (second reference): Secs. 4.4.

Intro to Probability and Logistic Regression

Reading Exercise: Introduction to Machine Learning, Ethem Alpaydin (first reference): Appendix A

Reading Exercise: Introduction to Machine Learning, Ethem Alpaydin (first reference): Sec 10.7.1

Naive Bayes Classifier

Naive Bayes' Model
Main assumption: Independence. Why?
Naive Bayes' Method for Categorical Inputs

Naive Bayes' Method for Continuous Inputs
Numerical Stability
Merits and Demerits

Naive_bayes_classifier

Reading Exercise: Machine Learning, Tom M Mitchell (second reference): Secs. 6.9-6.10.

Decision Tree

Decision Tree for Classification and Regression

Motivation for Decision Tree: Interpretability
Elements of Decision Tree: Root, Nodes, and Leaves
When to Split a Node?
Concept of Impurity -- Which feature to consider in a given node
Entropy Measure and other measures
Dealing with categorical as well as numerical features
Rule Extraction from Decision Trees
Regression Trees
Pruning to address overfitting

Decision_tree

Reading Exercise: Introduction to Machine Learning, Ethem Alpaydin (first reference): Chapter 9.

Programming Session 1 (for Group 1)

1. **Introduction to Python**

- Python: Fun to Use
- Intro to Google Collab
- Intro to Numpy: Matrices/Vectors and Mathematical Manipulations
- Intro to Pandas: Dataset structure and easy analysis
- Intro to Matplotlib: Plotting Data

2. **K Nearest Neighbor**
3. **Linear and Logistic Regression**
4. **Naive Bayes Classifier**
5. **Decision Tree**

Introduction to Python

Codes of Tutorials (with assignments)

Programming Assignment

Bias-Variance Tradeoff and its Solutions

1. Bias-variance Tradeoff

- Graphical Illustration
- Mathematical derivation

2. Addressing Bias-variance issues

- Underfitting and Overfitting in various models
- Training and Validation Sets
- K-fold Cross validation (to eliminate the drawback of the above)
- Lasso Regression and Sparsity (why no closed form solution exists?)
- Ridge Regression and its Closed form solution

- Bagging to address Overfitting
- Boosting to address Underfitting

3. MLE and MAP Models
    1. Importance of Gaussian Distribution in Machine Learning (Central Limit Theorem)
    2. Interpretation of MSE loss in Machine Learning
    3. Interpretation of Lasso Regression using MAP
    4. Interpretation of Ridge Regression using MAP

Bias-variance_tradeoff_slides

Bias-variance-tradeoff Derivation (handout)

Reading Exercise: Introduction to Machine Learning, Ethem Alpaydin (first reference): Secs 4.6-4.9

MLE and MAP

Reading Exercise :Machine Learning, Tom M Mitchell (second reference): Secs 5.4.1, 6.1-6.4

Class Test 1

Class Test 1 Solution

Class Test 1 Marks

Feed Forward Neural Networks

1. XOR problem: Solution with Multi-layer Neural Networks
2. Biological Motivation of Feedforward Neural Network
3. Machine Learning Vs Deep Learning

- Case study: Face detection
- Advantages of Deep Learning over Machine Learning
- Advantages of Machine Learning over Deep Learning

4. Feed Forward Neural network

- Overview
    - Terminologies (input, output, and hidden layers, and learnable parameters)
    - Impact of Bias (similar to the threshold in Perceptron)
    - Extracting low-level to high-level features, and Connectionism
- Forward Pass
    - Matrix multiplication based mapping
    - The need for Non-linearities between layers
- Backpropagation for optimizing weights
    - Vanishing/Exploding Gradient
    - Gradients for MSE, Linear Non-linear Activations, and Linear Layer
    - Reuse of component Gradients for different parameters

Feed_forward_NN

Reading Exercise: Machine Learning, Tom M Mitchell (second reference): Secs. 11.5, 11.7-11.9; 12.1-12.2.

Convolutional Neural Network

1. Biological Motivation
2. CNNs from MLPs
3. Different Layers of CNNs
4. Applications of CNNs for Classification and Regression
5. Standard CNN Architectures

- AlexNet

- VGGNet
- GoogleNet
- ResNet

CNN

Reading Exercise: Secs. 9.1 to 9.5 of https://www.deeplearningbook.org/contents/convnets.html.

Recurrent Neural Network

1. Feedforward vs Recurrent Networks
2. Motivation for RNNs
3. Different RNN types and Applications

- One-to-Many
- Many-to-One
- Many-to-Many

4. RNN model and parameters
5. Optimizing RNNs (Backpropagation through time)
6. Applications of RNNs

RNN

Reading Exercise: Secs. 10.1 to 10.2 of https://www.deeplearningbook.org/contents/rnn.html.

Mid-Semester Evaluation

Midsem Marks

Dimensionality Reduction Techniques

Role of Dimensionality Reduction in Machine Learning
Feature Selection Vs Feature Extraction
Feature Selection
Forward Search
Backward Search
Feature Extraction
Principal Component Analysis (PCA)
Linear Discriminant Analysis (LDA)

Dimensionality_reduction

Reading Exercise: Introduction to Machine Learning, Ethem Alpaydin (first reference): Secs 6.1-6.3 and 6.8

Support Vector Machines

Motivation of SVMs
Hard-margin Hyperplane SVMs
Soft-margin Hyperplane SVMs
Kernel SVMs
Optimization Problems for different SVMs

SVM overview

Reading Exercise: Intro to ML by Ethem Alpaydin (first reference): Secs. 13.1 - 13.5.

Ensemble Learning

Motivation of Ensemble Learning
Insights on How Bagging and Boosting reduce Variance and Bias, respectively.
Random Forests
Adaboost

Ensemble learning

Unsupervised Learning

Clustering -- Unsupervised Learning
Flat Clustering
K Nearest Neighbor --> Compactness
Spectral Clustering --> Connectivity
Mixture of Gaussian --> Soft Clustering
Hierarchical Clustering
Divisive -- > top-down
Agglomerative --> bottom-up

K Means Clustering

Hierarchical Clustering

Spectral Clustering and Mixture of Gaussian

Reinforcement Learning

Intro to Reinforcement Learning

Agent, Environment, Action, States
Mathematical formulation of Reinforcement Learning

Intro to Reinforcement Learning

Programming Session 2 (for Group 1)

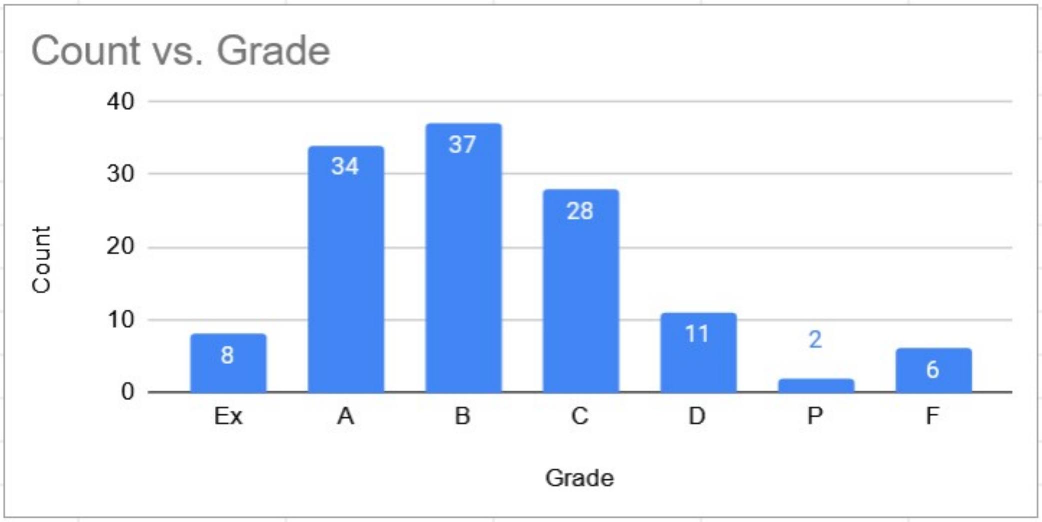
Intro to Pytorch

MLP and CNN

Class Test 2

Class Test 2 Marks

EndSem Evaluation and Course Grade Statistics



FYI: Six students who were absent in the Endsem exam got F grade (this is by default done by ERP).

Thanks to the students for helping me learn. Thanks to our TAs: Subhadip, Chhavi, and Nithya for their help and valuable time spent throughout the course.

We hope you learnt something from this course.

Good luck and Godspeed!

Endsem marks

Moodle Docs for this page

You are logged in as Dr. Mahesh Mohan MR (Log out)
Home

Data retention summary