Course Syllabus (download as pdf)

LECTURES:

I will provide course notes as well as slides for each lecture. Those will be uploaded to Canvas **before** the lecture. Feel free to print them and bring them to class with you for annotating.

You may also benefit from the recommended books (listed under Resources, see left tab) to further your understanding. To stay on track, make sure to read the course notes in a timely fashion, and follow up with questions in lectures, office hours, recitations, and/or Piazza.

RECITATIONS:

There will be a recitation session held by one of the TAs on Fridays 5:30-7pm. The recitation will review the week's material and answer any questions you might have about the course material, including homework.

Week | Lectures | Notes

Week 1

INTRO TO MACHINE LEARNING [+]

HW 0 out • Python and Jupyter setup

- The Learning Problem, Terminology
- Canonical Learning Problems
  - Supervised Learning
    - Regression
    - Classification (binary vs. multi-class)
  - Unsupervised Learning
    - Density estimation
    - Clustering
    - Dimensionality reduction
- ML applications in the real world
- What does it mean to learn?
  - A key ML concept: Generalization
  - vs. Overfitting
- Course Logistics

DATA PREPARATION [+]

Recitation 1 • Python setup • Data prep

- Python for ML Intro
- Feature Engineering
- Preliminary Data Analysis
  - EDA: exploratory data analysis
    - 1D: bar chart, histogram, box plot;
    - 2D: scatter plot, heat map and contourmap;
    - >3D: parallel coordinates, radar plot
- Data Cleaning and Transformation
  - Handling missing values
    - mean/median, kNN, model-driven imputation
  - Transforming feature types and feature values

- OHE: one-hot-encoding
- normalization
- log-transform

PART I: SUPERVISED LEARNING

Week 2 **LINEAR REGRESSION (LR)** [+]

- Formalizing the Learning Problem
  - loss functions
  - data generating distribution
  - models, parameters, hyperparameters
  - optimization algorithms
- Supervised Learning Cycle
- Linear models and Parameters
- Closed-form opt. for squared loss
- Interpreting coefficients
- Regularization
- Shrinkage methods: Ridge & Lasso regression
- Beyond linearity
  - Non-linear basis expansions
  - Local regression (*)
  - GAMs: Generalized Additive Models
- Practical issues:
  - feature scaling
  - categorical features, OHE
  - outliers & high-leverage points
  - collinearity
  - high dimensions

Recitation 2 Data prep demos • Linear Algebra review

Week 3 **MODEL SELECTION** [+]

- What is a good model?
- Overfitting and Generalization
- Decomposition of error
  - estimation vs. approximation error
- Bias-Variance tradeoff
- Regularization
- Separation of training and test data
- CV: Cross Validation

Recitation 3 • Linear Reg. demos • Convex optimization basics

HW 1 out • EDA • LR • Model selection • LogR

Week 4 **LOGISTIC REGRESSION (LogR)** [+]

- Classification vs. Regression
- 0-1 loss
- Convex surrogate loss functions & logistic loss
- Decision rule and boundary
- Intro to convex optimization basics
- Gradient descent optimization
- LR with >2 classes

Recitation 4 Bias-Variance trade-off • Cross-validation

- Kernel Logistic Regression (*)

NON-PARAMETRIC LEARNING [+]

Week 5 • Week 6

- k Nearest Neighbors (kNN) classifier
  - decision boundaries
- kNN regression
- Local regression
- Locally-weighted linear regression

- Comparison of LR/LogR with kNN

- Practical issues:
  - curse of dimensionality
  - intelligibility
  - computational efficiency
  - distance functions

Recitation 5 LogR • Gradient descent review and demos

HW 2 out • Non-parametric learning • Model evaluation • DT

Recitation 6 • kNN • Kernel regression • Model evaluation

MODEL EVALUATION [+]

- Evaluation metrics
  - Cost of false positives and false negatives
  - Confusion matrix
  - Visualizing model performance
    - ROC, precision-recall, lift, profit curves

- Debugging your model
  - train/test mismatch
  - analyzing error, ablative analysis
  - class imbalance and resampling strategies

- Creating baseline methods for comparison
- Statistical comparison of models

Week 7

DECISION TREES (DT) [+]

- Classification trees
- Regression trees
- Regularization and pruning

- Trees vs. Linear models

- Practical issues:
  - handling missing values

Recitation 7 • DT review and demos

Week 8

Midterm Review

Midterm Exam

Exam will be during class on Thur. Duration: 80 minutes. You can only bring your own notes up to 2 A4-size sheets. No electronics.

Friday NO RECITATION

Week 9

NO CLASS: Spring Break

Week 10

ENSEMBLE METHODS [+]

- Combining multiple models
- Bagging
- Random Forests
- Boosting

HW 3 out • Ensembles • NB • SVM

Recitation 8 Random Forest • Boosting • NB

Case Study out • Dataset provided, Tasks recommended

NAIVE BAYES (NB) [+]

- Classification by density estimation
- Conditional independence
- MLE, Regularization via priors and MAP
- Generative vs. Discriminative models
- Gaussian NB (*)

Week 11

SUPPORT VECTOR MACHINES (SVM) [+]

- SVM formulation
  - construction of the max-margin classifier
- The non-separable case
  - hard vs. soft-margin SVM
  - slack variables
- Hinge loss
- SVMs with >2 classes
- Relation to LR
- Intro to dual optimization
- SVM dual
- The Kernel trick
  - From feature combinations to kernels
  - Kernel SVM
  - Interpreting SVM dual and its solution
  - (*) Kernel Logistic Regression

Recitation 9 • SVM and Kernels

Week 12

•

NEURAL NETWORKS (NN) [+]

- Representation
  - Perceptron
  - single- & multi-layer networks
  - multiclass classification
- Learning
  - Backpropagation algorithm
  - Regularization

HW 4 out • Kernels • Neural Nets • Density estimation

Recitation 10 • NNs • Back-propagation

Week 13

PART II: UNSUPERVISED LEARNING

DENSITY ESTIMATION [+]

- Parametric
  - Gaussian/Poisson/etc

- Gaussian/Poisson/etc.
    - MLE: Maximum Likelihood Estimation
    - MAP: Maximum A Posteriori estimation

- Non-parametric
    - Histograms
    - KDE: Kernel Density Estimation

Thur NO CLASS: Spring carnival

Friday NO RECITATION

Week 14 • Week 15

CLUSTERING [+]

- Similarity/distance functions
- Hierarchical clustering
- k-means clustering
- Mixture models
- EM: Expectation Maximization

DIMENSIONALITY REDUCTION [+]

- Unsupervised embedding techniques
    - PCA: Principal Component Analysis
    - Kernel PCA
    - t-SNE
    - MDS: Multi-Dimensional Scaling
- Supervised reduction techniques
    - Feature selection
        - forward selection
        - backward selection

HW 5 out • Clustering • EM • Dimensionality reduction

Recitation 11 • Density estimation • hierarchical clustering • k-means

Recitation 12 • EM • Dim. reduction

Week 16

Case Study & Final Review

Recitation 13 • Case Study review • Final Q&A

Last modified by Leman Akoglu, 2019