JTL 312: Introduction to Machine Learning (Jan-May, 2025)

Instructor: Rajdeep Banerjee, PhD
Teaching Assistant: Nimish Dwarkanath, PhD

Syllabus:

1. Introduction to probability and information theory (week 1, theory: 2 Hr, hands-on: 2 Hr)
    - 1.1. Definitions of probability
    - 1.2. Random variables
    - 1.3. Probability distributions – marginal, conditional, chain rule
        - 1.3.1. Binomial distribution
        - 1.3.2. Normal distribution
        - 1.3.3. Student's t-distribution
        - 1.3.4. F-distribution
    - 1.4. Sample/Population metrics
        - 1.4.1. Expectations
        - 1.4.2. variance
        - 1.4.3. skew
        - 1.4.4. covariance
        - 1.4.5. Anscombe's quartet
    - 1.5. Discrete vs. continuous probability distributions
    - 1.6. Fundamentals of sampling distributions – central limit theorem
    - 1.7. Hypothesis testing
        - 1.7.1. ANOVA
        - 1.7.2. Student's t-test
        - 1.7.3. Chi-squared test

Lab session:

- 1.8. Python basics
    - 1.8.1. Data structures, loops, functions
    - 1.8.2. Pandas, NumPy
    - 1.8.3. OOP
    - 1.8.4. Statistics with Python

2. Bayes' rule, Bayesian networks (week 2, theory: 2 Hr, hands-on: 2 Hr)
    - 2.1. Bayes' revisit - probabilistic graphical model
    - 2.2. MLE
    - 2.3. Information theory – entropy, mutual information, KL-divergence
    - 2.4. Basics of some probabilistic concepts used in machine learning (conceptual introduction)
        - 2.4.1. Naïve Bayes (theory)

Lab session: Exploratory data analysis with Python

- 2.5. Data pre-processing
- 2.6. Handling missing values
- 2.7. Handling categorical variables
- 2.8. Feature engineering
- 2.9. Feature selection

2.10. Scaling
2.11. Univariate and bivariate analysis
2.12. Naive Bayes (lab)

2. Introduction to machine learning (week 3, theory: 2 Hr, hands-on: 2 Hr)
    2.1. Statistical learning
        2.1.1. Supervised vs. unsupervised
        2.1.2. Regression vs. classification
        2.1.3. Empirical risk minimization and prior knowledge (no-free-lunch theorem)
    2.2. Accuracy vs. interpretability
    2.3. Hyperparameters, train-validation-test splits
    2.4. Bias-variance trade-off

3. Regression (week 4, theory: 2 Hr, hands-on: 2 Hr)
    3.1. Linear regression
        3.1.1. Problem Statement
        3.1.2. Loss/cost function and MLE
    3.2. Estimating coefficients
        3.2.1. Exact Solution
        3.2.2. Regression as an optimization problem
            3.2.2.1. Gradient Descent (GD)
            3.2.2.2. Stochastic GD
            3.2.2.3. Mini-Batch GD
    3.3. Estimating errors
        3.3.1. R^2
        3.3.2. Adjusted R^2
        3.3.3. MAPE, MAE, etc. (when to use what)
        3.3.4. AIC, BIC
    3.4. Hypothesis Testing
        3.4.1. What's an F-score for a regression fit?

    3.5. Shrinkage methods (Regularization)
        3.5.1. Ridge regression
        3.5.2. Lasso regression
        3.5.3. Regularization as a consequence to MAP/Bayesian inference
        3.5.4. Ridge as gaussian prior
        3.5.5. Lasso as Laplace prior
    3.6. Lab session
        3.6.1. Introduce Statsmodels for regression

4. Classification (week 5, theory: 2 Hr, hands-on: 2 Hr)
    4.1. Logistic regression
        4.1.1. Estimating probabilities
        4.1.2. Cost function
        4.1.3. Choosing the probability threshold

    4.2. K-nearest neighbours
    4.3. Naïve Bayes (revisit)
    4.4. Linear discriminant analysis
    4.5. Comparing different classification models (Lab)

5. Support vector machines (week 6, theory: 2 Hr, hands-on: 2 Hr)
    5.1. Maximal margin classifier
    5.2. Support vector classifier
        5.2.1. Linear- Soft-margin
        5.2.2. Non-linear – the kernel trick
    5.3. Support vector regressor
    5.4. Relation to logistic regression
    5.5. Lab session

6. Unsupervised learning – dimensionality reduction and clustering (week 7, theory: 2 Hr, hands-on: 2 Hr)
    6.1. The curse of dimensionality
    6.2. Principal component analysis
    6.3. K-means clustering
    6.4. Hierarchical clustering
    6.5. Gaussian mixtures (if time permits)
    6.6. Lab session

7. Decision trees (week 8, theory: 2 Hr, hands-on: 2 Hr)
    7.1. Regression trees
    7.2. Classification trees
    7.3. The CART algorithm
    7.4. Computational complexity
    7.5. Gini or entropy?
    7.6. Lab session

8. Ensemble learning – bagging (week 9, theory: 2 Hr, hands-on: 2 Hr)
    8.1. Bagging and pasting
    8.2. Out of bag evaluation
    8.3. Random forests – why?
    8.4. Feature importance
    8.5. Lab session

9. Ensemble learning – boosting methods (week 10, theory: 2 Hr, hands-on: 2 Hr)
    9.1. Boosting
    9.2. AdaBoost
    9.3. Gradient boost
    9.4. XGBoost (introduction)
    9.5. Comparison of performance (Lab)

10. Model analysis and data selection (week 11, theory: 2 Hr, hands-on: 2 Hr)

10.1. Cross-validation
    10.1.1. Leave-one-out (LOOCV)
    10.1.2. k-fold
10.2. Bootstrap
10.3. Introduction to data-centric AI
    10.3.1. Detecting label issues
    10.3.2. Data selection for retraining
10.4. Lab session

11. Back to Bayesian (week 12, theory: 2 Hr, hands-on: 2 Hr)
    11.1. Conjugate priors
        11.1.1. example: beta and binomial distributions
    11.2. Expectation maximization and Gausian mixture models
    11.3. Variational inference - mean-field approximation
    11.4. Markov chain Monte Carlo
    11.5. Lab session

12. (If time permits) Introduction to neural networks (week 13, theory: 2 Hr, hands-on: 2 Hr)
    12.1. The perceptron
    12.2. Building a neural network from scratch
        12.2.1. The feed-forward neural network
        12.2.2. Back-propagation
    12.3. Stop overfitting
        12.3.1. Drop-outs
        12.3.2. Early-stop
        12.3.3. Batch normalization
    12.4. The PyTorch and Tensorflow frameworks (Lab)

13. Project/paper presentation, Q&A (week 14)