SYLLABUS

Basic math: vectors, subspaces, matrices, rank, pseudo inverse, tensor transpose, multi-variate function gradient, gradient descent, Lagrangian for constrained optimization, discrete random variable, PMF, entropy, joint, conditional, marginal, continuous RV, PDF, Gaussian, uniform, exponential, maximum likelihood parameter estimation, multivariate distribution basic concepts, multivariate Gaussian, uniform. **Bishop's book: Appendices B, C, E (only topics mentioned above), Chapter Sections 1.1, 1.2, 1.2.1-1.2.4, 2.3, 2.3.1, 2.3.2**

EDA: Data types, conversion between data types (discrete to integer, integer to dummies, integer to float, float to integer, text to integer), histograms, bin size in histograms, descriptive statistics (mean, median, percentiles, max, min, std dev), missing values and NaNs, correlation, scatter plots.

Basic statistical testing: IID assumption, data log likelihood, comparing distributions for given data, confidence interval around a mean, independent t-test (ignore DoF), paired t-test (ignore DoF), Wilcoxon rank-sum test, Wilcoxon signed-rank test, Pearson's correlation, Spearman's correlation.

Graphs: Bar, line, pie, scatter, stacked bar, stacked bar 100%, box and whiskers, QQ, data transformations (log, polar), improving plots.

SQL: Basics of select, from, where, join

Intro to ML: Training, validation, testing, parameter, hyper-parameter, supervised learning -- classification vs regression, unsupervised learning -- clustering vs dimension reduction, overfitting vs underfitting concept

Linear regression: Basic model, noise assumption, derivation of MSE as loss function from probabilistic assumptions, pseudoinverse times t as the maximum likelihood solution, extension to MAE for Laplace error distribution, bias-variance-noise decomposition, form of L2 and L1 penalty, geometry of L2 and L1 regularization (including variable elimination for L1), derivation of gradient descent for L2 vs L1. **Bishop's book Sections 3, 3.1, 3.1.1, 3.1.2, 3.1.4, 3.2**

Linear classification: Basic decision criteria form, derivation of linear classifier as Bayesian optimal for Gaussian class conditionals with same covariance matrices, basic recipe for building a Bayesian classifier with extension to non-Gaussian class conditionals, logistic regression and its gradient descent, L2 regularization. **Bishop's book Sections 4.2, 4.2.1, 4.2.2, 4.3.2**.

Additional: elastic net and binary classification Metrics -- FP, FN, TP, TN, accuracy, sensitivity, specificity, precision, recall, F1, AUC-ROC, and also the concept of asymmetric risk.