DATA 1010 Course Standards

Students will be able to...

SETFUN Correctly answer questions about basic set and function terminology

JULIA Write Julia code to solve simple algorithmic problems using conditionals, functions, arrays, dictionaries, and iteration.

LINALG Use vocabulary and results from linear algebra to solve problems involving linear independence, span, and rank.

MATALG Use matrix algebra (including matrix transposes) to solve problems involving projection and orthogonality

EIGEN Apply knowledge of determinants, eigendecomposition, and singular value decomposition to data problems and other applications

OPT Explain the Lagrange multipliers theorem and gradient descent and discuss issues surrounding applied optimization

MATDIFF Differentiate matrix expressions with respect to vectors and use this technique to solve optimization problems.

MACHARITH Reason about 64-bit and 32-bit floating point arithmetic

NUMERROR Discuss the categories of numerical error and identify points of concern in application

PRNG Discuss basic considerations surrounding the generation of pseudorandom numbers, such as seed, period, and statistical tests

COUNTING Use the fundamental principle of counting and binomial coefficients to solve basic counting problems

PROBSPACE Explain the elements of a probability space and use probability spaces to model random experiments

PMF Reason about discrete random variable distributions and use properties of discrete distributions to solve problems

PDF Reason about continuous random variable distributions and use properties of continuous distributions to solve problems

CONDPROB Use the conditional probability formula to translate back and forth between branching tree diagrams and their corresponding probability spaces

BAYES Use Bayes' theorem and other properties of conditional probability to solve conditional probability problems

IND Explain independence of random variables, construct a probability space with independent random variables, and use independence to solve probability problems

EXP Use the definition of a random variable, the distribution of the random variable, or linearity of expectation to find the expectation of a random variable

COV Calculate variances and covariances, recognize high or low variance and positive or negative covariance from graphical representations of distributions, and use properties of variance and covariance to solve problems about random variable distributions

CONDEXP Calculate conditional expectations and apply them to expectation problems

COMDISTD Discuss definitions and properties of common discrete distributions (Bernoulli, binomial, geometric, Poisson) and recognize circumstances under which those distributions can be expected to fit the data well

COMDISTC Discuss definitions and properties of common continuous distributions (exponential, uniform, multivariate normal)

RVINEQ Explain inequalities involving random variable expectations (such a Chebyshev's inequality) and use them to solve problems

CLT State and apply the central limit theorem, and recognize when the conclusion of the central limit theorem should not be expected to hold

KDE Apply kernel density estimators to data problems, and explain ways of dealing with the bias-variance tradeoff in density estimation

LR Explain the techniques of basic linear and polynomial regression, and discuss the advantages and disadvantages relative to nonparametric methods

QDA Discuss the assumptions of, the estimation methods for, and facts about quadratic and linear discriminant analysis

STATLEARN Explain the main points of statistical learning theory (regression vs classification, loss functional, target function, learner, training and test error, overfitting, inductive bias, bias-variance tradeoff)

NPL Apply classification vocabulary (confusion matrix, detection rate, false alarm rate, precision, receiver operating characteristic) and the Neyman-Pearson lemma to reason about classification problems

SVM Describe the mathematics and intuition behind support vector machines (both hard- and soft-margin)

LOGIST Describe, apply, and analyze logistic regression models

NN Describe, apply, and analyze multi-layer perceptrons for regression and classification

DR Describe and interpret dimension reduction methods, including principal component analysis (concept and technical details) and t-SNE (concept only)

R Perform basic programming tasks in R (defining variables, generating and indexing matrices, control flow, and writing functions)

GGPLOT Use `ggplot` to create data visualizations (data, aesthetics, geometries, statistics, scales, faceting)

DPLYR Apply the six fundamental verbs in Hadley Wickham's grammar of data manipulation (`filter`, `arrange`, `select`, `mutate`, `group_by`, `summarise`) to transform data

POINTEST Discuss the relationship between bias and consistency, determine whether a given estimator is biased or consistent, and calculate and interpret confidence intervals

HYPTEST Perform a hypothesis test and interpret hypothesis test findings (including multiple hypothesis testing)

BOOT Apply the Glivenko-Cantelli theorem and use the bootstrap method to estimate statistical functionals

MLE Calculate maximum likelihood estimators, and give examples to illustrate the shortcomings of MLE