DATA 1010 Course Standards

Students will be able to...

JULIA	1	Write Julia code to solve simple algorithmic problems using conditionals, functions, arrays, dictionaries, and iteration.
LINALG	2	Use vocabulary and results from linear algebra to solve problems involving linear independence, span, and rank.
MATALG	3	Use matrix algebra (including matrix transposes) to solve problems involving projection and orthogonality
EIGEN	4	Apply knowledge of determinants, eigendecomposition, and singular value decomposition to data problems and other applications
OPT	5	Solve problems using the Lagrange multipliers theorem and/or the relationship between critical points of multivariable functions and Hessian eigenvalues
MATDIFF	6	Differentiate matrix expressions with respect to vectors and use this technique to solve optimization problems.
MACHARITH	7	Reason about 64-bit and 32-bit integer and floating point arithmetic
NUMERROR	8	Discuss the categories of numerical error and identify points of concern in applications
PRNG	9	Discuss basic considerations surrounding the generation of pseudorandom numbers, such as seed, period, and statistical tests
NUMOPT	10	Reason about gradient descent algorithms and discuss issues surrounding applied optimization
PROBSPACE	11	Explain the elements of a probability space and use probability spaces to model random experiments
CONDPROB	12	Use the conditional probability formula to translate back and forth between branching tree diagrams and their corresponding probability spaces
BAYES	13	Use Bayes' theorem and other properties of conditional probability to solve conditional probability problems
IND	14	Explain independence of random variables, construct a probability space with independent random variables, and use independence to solve probability problems
EXP	15	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	16	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	17	Calculate conditional expectations and conditional variances and apply them to expectation problems
COMDISTD	18	Discuss definitions and properties of common discrete distributions (Bernoulli, binomial, geometric, Poisson) and recognize circumstances under which those distributions can be expected to fit observed data well
COMDISTC	19	Discuss definitions and properties of common continuous distributions (exponential, uniform, multivariate normal)
CLT	20	State and apply the central limit theorem, and recognize when the conclusion of the central limit theorem should not be expected to hold
POINTEST	21	Discuss the relationship between bias and consistency, determine whether a given estimator is biased or consistent, and calculate and interpret confidence intervals
BOOT	22	Apply the Glivenko-Cantelli theorem and use the bootstrap method to estimate statistical functionals
HYPTEST	23	Perform a hypothesis test and interpret hypothesis test findings (including multiple hypothesis testing)
MLE	24	Calculate maximum likelihood estimators, and give examples to illustrate the shortcomings of MLE
STATLEARN	25	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)
LRC	26	Apply classification vocabulary (confusion matrix, detection rate, false alarm rate, precision, receiver operating characteristic) and the Neyman-Pearson lemma to reason about classification problems
KDE	27	Apply kernel density estimators to data problems, and explain ways of dealing with the bias-variance tradeoff in density estimation
LR	28	Explain the techniques of basic linear and polynomial regression, and discuss the advantages and disadvantages relative to nonparametric methods
LOGIST	29	Describe, apply, and analyze logistic regression models
QDA	30	Discuss the assumptions of, the estimation methods for, and facts about quadratic and linear discriminant analysis
SVM	31	Describe the mathematics and intuition behind support vector machines (both hard- and soft-margin, and SVM with radial basis function kernel)
DT	32	Train and interpret decision trees for classification and regression.
ENSEMBLE	33	Discuss the mechanisms behind and benefits of common ensemble methods, including bagging and gradient boosting.
DR	34	Describe and interpret dimension reduction methods, including principal component analysis and t-SNE
NN	35	Describe, apply, and analyze multi-layer perceptrons for regression and classification
FREQBAYES	36	Describe the distinction between frequentist and Bayesian approaches to statistics, and explain computational methods to Bayesian analysis
MCMC	37	Solve exercises about basic Markov chain Monte Carlo theory, and discuss how MCMC relates to graphical models and to Bayesian methods
GM	38	Use graphical models (including Gaussian mixture models and hidden Markov models) to represent relationships between random variables in applications
PP	39	Explain the purpose of probabilistic programming and and use a probabilistic programming system to perform inference
R	40	Perform basic programming tasks in R (defining variables, generating and indexing matrices, control flow, and writing functions), use `ggplot` to create data visualizations (data, aesthetics, geometries, statistics, scales, faceting), and apply the six fundamental verbs in Hadley Wickham's grammar of data manipulation (`filter`, `arrange`, `select`, `mutate`, `group_by`, `summarise`) to transform data