GATE DA (Data Science and Artificial Intelligence) – Syllabus 2025

Probability and Statistics:

Counting (permutation and combinations),probability axioms, Sample space, events, independent events, mutually exclusive events, marginal, conditional and joint probability, Bayes Theorem, conditional expectation and variance, mean, median, mode and standard deviation, correlation, and covariance, random variables, discrete random variables and probability mass functions, uniform, Bernoulli, binomial distribution, Continuous random variables and probability distribution function, uniform, exponential, Poisson, normal, standard normal, t-distribution, chi-squared distributions, cumulative distribution function, Conditional PDF, Central limit theorem, confidence interval, z-test, t-test, chi-squared test.

Linear Algebra:

Vector space, subspaces, linear dependence and independence of vectors, matrices, projection matrix, orthogonal matrix, idempotent matrix, partition matrix and their properties, quadratic forms, systems of linear equations and solutions; Gaussian elimination, eigenvalues and eigenvectors, determinant, rank, nullity, projections, LU decomposition, singular value decomposition. Calculus and Optimization: Functions of a single variable, limit, continuity and differentiability, Taylor series, maxima and minima, optimization involving a single variable.

Programming, Data Structures and Algorithms:

Programming in Python, basic data structures: stacks, queues, linked lists, trees, hash tables; Search algorithms: linear search and binary search, Sorting algorithms: selection sort, bubble sort and insertion sort; divide and conquer: merge sort, quicksort; introduction to graph theory; basic graph algorithms: traversals and shortest path.

Database Management and Warehousing:

ER-model, relational model: relational algebra, tuple calculus, SQL, integrity constraints, normal form, file organization, indexing, data types, data transformation such as normalization, discretization, sampling, compression; data warehouse modelling: schema for multidimensional data models, concept hierarchies, measures: categorization and computations.

Machine Learning:

(i)Supervised Learning:

regression and classification problems, simple linear regression, multiple linear regression, ridge regression, logistic regression, k-nearest neighbour, naive Bayes classifier, linear discriminant analysis, support vector machine, decision trees, bias variance trade-off, cross-validation methods such as leave-one-out (LOO) cross-validation, k-folds cross-validation, multi-layer perceptron, feed-forward neural network;

(ii) Unsupervised Learning:

Clustering algorithms, k-means/k-medoid, hierarchical clustering, top-down, bottom-up: single linkage, multiple-linkage, dimensionality reduction, principal component analysis.

Artificial Intelligence:

Search: informed, uninformed, adversarial; logic, propositional, predicate; reasoning under uncertainty topics — conditional independence representation, exact inference through variable elimination, and approximate inference through sampling.