GLOSSARY FOR ISYE 6501 INTRODUCTION TO ANALYTICS MODELING

(Organized by topics; for full alphabetical glossary, see other file)

BASIC MACHINE LEARNING	***LESSONS 2.1-2.2, 2.4-2.6, 2.8, 4.1, 4.3-4.6, 6.1-6.3, 16.4***
Algorithm	Step-by-step procedure designed to carry out a task.
Change detection	Identifying when a significant change has taken place in a process.
Classification	The separation of data into two or more categories, or (a point's classification) the category a data point is put into.
Classifier	A boundary that separates the data into two or more categories. Also (more generally) an algorithm that performs classification.
Cluster	A group of points identified as near/similar to each other.
Cluster center	In some clustering algorithms (like k-means clustering), the central point (often the centroid) of a cluster of data points.
Clustering	Separation of data points into groups ("clusters") based on nearness/similarity to each other. A common form of unsupervised learning.
CUSUM	Change detection method that compares observed distribution mean with a threshold level of change. Short for "cumulative sum".
Deep learning	Neural network-type model with many hidden layers.
Dimension	A feature of the data points (for example, height or credit score). (Note that there is also a mathematical definition for this word.)
EM algorithm	Expectation-maximization algorithm.
Expectation-maximization algorithm (EM algorithm)	General description of an algorithm with two steps (often iterated), one that finds the function for the expected likelihood of getting the response given current parameters, and one that finds new parameter values to maximize that probability.
Heuristic	Algorithm that is not guaranteed to find the absolute best (optimal) solution.
k-means algorithm	Clustering algorithm that defines k clusters of data points, each corresponding to one of k cluster centers selected by the algorithm.
k-Nearest-Neighbor (KNN)	Classification algorithm that defines a data point's category as a function of the nearest k data points to it.
Kernel	A type of function that computes the similarity between two inputs; thanks to what's (really!) sometimes known as the "kernel trick", nonlinear classifiers can be found almost as easily as linear ones.
Learning	Finding/discovering patterns (or rules) in data, often that can be applied to new data.
Machine	Apparatus that can do something; in "machine learning", it often refers

	to both an algorithm and the computer it's run on. (Fun fact: before computers were developed, the term "computers" referred to people who did calculations quickly in their heads or on paper!)
Margin	For a single point, the distance between the point and the classification boundary; for a set of points, the minimum distance between a point in the set and the classification boundary. Also called the separation.
Machine learning	Use of computer algorithms to learn and discover patterns or structure in data, without being programmed specifically for them.
Misclassified	Put into the wrong category by a classifier.
Neural network	A machine learning model that itself is modeled after the workings of neurons in the brain.
Supervised learning	Machine learning where the "correct" answer is known for each data point in the training set.
Support vector	In SVM models, the closest point to the classifier, among those in a category. (Note that there is a more-technical mathematical definition too.)
Support vector machine (SVM)	Classification algorithm that uses a boundary to separate the data into two or more categories ("classes").
SVM	Support vector machine.
Unsupervised learning	Machine learning where the "correct" answer is not known for the data points in the training set.
Voronoi diagram	Graphical representation of splitting a plane with two or more special points into regions with one special point each, where each region's points are closer to the region's special point than to any other special point.

CONFUSION MATRICES	***LESSONS 10.5-10.6***
Accuracy	Fraction of data points correctly classified by a model; equal to $\frac{TP+TN}{TP+FP+TN+FN}$.
Confusion matrix	Visualization of classification model performance.
Diagnostic odds ratio	Ratio of the odds that a data point in a certain category is correctly classified by a model, to the odds that a data point not in that category is incorrectly classified by the model; equal to $\frac{TP/FN}{FP/TN} = \frac{TN \times TP}{FN \times FP}$.
Fall out	Fraction of data points not in a certain category that are incorrectly

	classified by a model; equal to $\frac{FP}{TN+FP}$. Also called false positive rate.
False negative (FN)	Data point that a model incorrectly classifies as not being in a certain category. ("Negative" means the model classified it as not being in the category, and "False" means the model's classification is incorrect.) Sometimes abbreviated as "FN".
False negative rate	Fraction of data points in a certain category that are incorrectly classified by a model; equal to $\frac{FN}{TP+FN}$. Also called miss rate.
False positive (FP)	Data point that a model incorrectly classifies as being in a certain category. ("Positive" means the model classified it as being in the category, and "False" means the model's classification is incorrect.) Sometimes abbreviated as "FP".
False positive rate	Fraction of data points not in a certain category that are incorrectly classified by a model; equal to $\frac{FP}{TN+FP}$. Also called fall out.
False omission rate	Fraction of data points the model classifies as not in a certain category, that are really in the category; equal to $\frac{FN}{TN+FN}$.
Hit rate	Fraction of data points in a certain category that are correctly classified by a model; equal to $\frac{TP}{TP+FN}$; also called the true positive rate, sensitivity, and recall.
Miss rate	Fraction of data points in a certain category that are incorrectly classified by a model; equal to $\frac{FN}{TP+FN}$. Also called false negative rate.
Negative likelihood ratio	Ratio of the fraction of data points in a certain category that are misclassified as not in the cateogry, to the fraction of data points not in the category that are correctly classified as not being in the category; equal to (1-sensitivity)/specificity = $\frac{FN/(FN+TP)}{TN/(TN+FP)}$.
Negative predictive value	Fraction of data points classified as not in a certain category that are really not in that category; equal to $\frac{TN}{TN+FN}$.
Positive likelihood ratio	Ratio of the fraction of data points in a certain category that are correctly classified as being in that category, to the fraction of data points not in the category that are incorrectly classified as being in the category; equal to sensitivity/(1-specificity) = $\frac{TP/(TP+FN)}{FP/(FP+TN)}$.
Positive predictive value	Fraction of data points classified as being in a certain category that are really in that category; equal to $\frac{TP}{TP+FP}$. Also called precision.
Precision	In analytics, the fraction of data points classified as being in a certain

	category that are really in that category; equal to $\frac{TP}{TP+FP}$. Also called positive predictive value.
Recall	Fraction of data points in a certain category that are correctly classified by a model; equal to $\frac{TP}{TP+FN}$; also called sensitivity, hit rate, and true positive rate.
Sensitivity	Fraction of data points in a certain category that are correctly classified by a model; equal to $\frac{TP}{TP+FN}$; also called the true positive rate, hit rate, and recall.
Specificity	Fraction of data points not in a certain category that are correctly classified by a model; equal to $\frac{TN}{TN+FP}$; also called the true negative rate.
True negative (TN)	Data point that a model correctly classifies as not being in a certain category. ("Negative" means the model classified it as not being in the category, and "True" means the model's classification is correct.) Sometimes abbreviated as "TN".
True negative rate	Fraction of data points not in a certain category that are correctly classified by a model; equal to $\frac{TN}{TN+FP}$; also called specificity.
True positive (TP)	Data point that a model correctly classifies as being in a certain category. ("Positive" means the model classified it as being in the category, and "True" means the model's classification is correct.) Sometimes abbreviated as "TP".
True positive rate	Fraction of data points in a certain category that are correctly classified by a model; equal to $\frac{TP}{TP+FN}$; also called sensitivity, hit rate, and recall.

DATA	***LESSONS 2.3, 2.7, 5.1-5.3, 9.2-9.5, 14.1-14.3***
Attribute	A characteristic or measurement – for example, a person's height or the color of a car. Generally interchangeable with "feature", and often with "covariate" or "predictor". In the standard tabular format, a column of data.
Binary data	Data that can take only two different values (true/false, 0/1, black/white, on/off, etc.).
Box and whisker plot	Graphical representation data showing the middle range of data (the "box"), reasonable ranges of variability ("whiskers"), and points (possible outliers) outside those ranges.
Categorical data	Data that classifies observations without quantitative meaning (for

	example, colors of cars) or where quantitative amounts are categorized (for example, "0-10, 11-20, …").
Collective outlier	A set of data points that is (uncommonly) different from others – for example, a missing heartbeat in an electrocardiogram; we don't know exactly which millisecond it should've happened in, but collectively there's a set of milliseconds that it's missing from.
Contextual outlier	A data point that is (uncommonly) far from other data points related to it – for example, in Atlanta, a 90-degree (Fahrenheit) day in winter is an outlier, but a 90-degree day in summer is not.
Covariate	A characteristic or measurement that can be used to estimate the value of something – for example, a person's height or the color of a car. A "feature" or "attribute"; in the standard tabular format, a column of data.
Data point	Observation/record of (perhaps multiple) measurements for a single member of a population or data set. In the standard tabular format, a row of data.
Detrending	Removal of trend, such as a change in the mean over time, from time-series data.
Eigenvalue	Amount by which an eigenvector gets rescaled in a linear transformation.
Eigenvector	Non-zero vector that does not change direction when a linear transformation is applied to it, but only gets rescaled by the eigenvalue
Feature	(1) A characteristic or measurement – for example, a person's height or the color of a car. Generally interchangeable with "attribute", and often with "covariate" or "predictor". In the standard tabular format, a column of data. Also called an attribute. (2) A combination of attributes in a specific format – for example, 0.5×height plus 7×shoe-size.
Imputation	Inserting values where data is missing.
Observation	(1) A measurement of one attribute of a data point. (2) A measurement of all attributes of a data point (i.e., a full row of data). (3) The act of watching/measuring/recording something.
Outcome	A variable of interest that a model tries to estimate or predict.
PCA	Principal component analysis.
Point outlier	A data point that is (uncommonly) far from other data points – for example, an outdoor temperature reading of 200 degrees Fahrenheit.
Predictor	A characteristic or measurement that is used to estimate ("predict")

	the future value of something – for example, a person's height or the color of a car. A "feature" or "attribute"; in the standard tabular format, a column of data.
Principal component analysis (PCA)	Transformation of data into orthogonal dimensions that are ranked by variance.
Quantitative data	Data that describes numerical amounts of something – for example, height and weight.
Response	A variable of interest that a model tries to estimate or predict.
Scaling	Shrinking or expanding, and moving, the range of data to fit exactly into a specific interval (for example, between 0 and 1, or between 100 and 800).
Standardization	Transforming data by subtracting the mean and then dividing by standard deviation, so that it has mean 0 and variance 1.
Structured data	Data that is highly organized, so it can be searched, queried, and analyzed easily – for example, a table with the name, age, and country of participants in this course.
Time series data	Data that records the same attribute/response at multiple points in time (often at equal time intervals).
Unstructured data	Data that is not very well organized for analysis – for example, a list of free responses to the question "What do you like about analytics?"

DESIGN OF EXPERIMENTS	***LESSONS 12.1-12.4***
A/B testing	Test of two alternatives to see if either one leads to better outcomes.
Analysis of Variance/ANOVA	Statistical method for dividing the variation in observations among different sources.
Balanced design	Set of combinations of factor values across multiple factors, that has the same number of runs for all combinations of levels of one or more factors.
Blocking	Factor introduced to an experimental design that interacts with the effect of the factors to be studied. The effect of the factors is studied within the same level (block) of the blocking factor.
Control	(1) A variable whose value remains constant for all runs of an experiment, so changes in this variable don't affect the experiment. (2) Design an experiment where some factors ("controls" by definition (1)) are held constant to avoid them affecting the outcome.
Design of experiments	Choosing a set of tests to be made to find the effect of input variables

	on an outcome.
Exploitation	Using known information to get good outcomes.
Exploration	Finding new/better/more information to determine how to optimize output.
Factorial design	Tests of different combinations of factor values over multiple factors, to find each one's effect, and interaction effects, on the outcome.
Fractional factorial design	Test of a subset of all possible combinations of factor values over multiple factors. If chosen well, the desired effects of factors and factor interaction effects can be obtained.
Full factorial design	Test of all possible combinations of factor values over multiple factors to find each one's effect, and interaction effects, on the outcome.
Multi-armed bandit	Model that allows the tradeoff between exploration of unknown resources and exploitation of known resources to optimize output.
Response surface	Sequential experimentation strategy to understand the relationship between response and input factors, and/or optimize the response.

GAME THEORY	***LESSONS 16.5-16.5a***
Cooperative game theory	A game theory setting where the participants are also working together to achieve some goal, while also competing in some way.
Game theory	The study of competitive strategic decision-making where the outcome of each participant's actions is dependent on another participant's actions.
Mixed strategy/randomized strategy	A strategy where a participant's action is determined randomly according to probabilities – for example, in "rock, paper, scissors", someone who randomly chooses between the three options with probability $^1/_3$ each is using a mixed strategy.
Prisoner's dilemma	A situation in game theory where each participant would benefit if all participants act in a certain way, but each participant individually has incentive to not act that way.
Pure strategy	A strategy where a participant's action is deterministic (known with probability 1) – for example, in "rock, paper, scissors", someone who always chooses "rock" is using a pure strategy.
Sequential game	A game in which participants choose their actions one after another, so participants who choose later have knowledge of the earlier actions.
Simultaneous game	A game in which all participants choose their actions at the same time.

Stable equilibrium	A situation in game theory where, given each participant's current choice of action, no participant can do better by changing actions.
Zero-sum game	A game where the total gain and loss of all participants is zero. Some participants might benefit and others might lose, but the total of all benefits is equal to the total of all losses.

MODEL QUALITY	***LESSONS 3.1-3.4, 8.2, 8.4***
AIC	Akaike information criterion
Akaike information criterion (AIC)	Model selection technique that trades off between model fit and model complexity. When comparing models, the model with lower AIC is preferred. Generally penalizes complexity less than BIC.
Bayesian Information criterion (BIC)	Model selection technique that trades off model fit and model complexity. When comparing models, the model with lower BIC is preferred. Generally penalizes complexity more than AIC.
BIC	Bayesian information criterion
Causation	Relationship in which one thing makes another happen (i.e., one thing causes another).
Corrected AIC	Improved version of AIC, especially when sample size is small.
Correlation	Relationship in which two things are likely to happen together, regardless of whether one causes the other. (There is also a quantitative statistical definition measuring the amount of correlation.)
Cross-validation	Validation technique where a model is tested on data different from what it was trained on.
Hypothesis test	Statistical test to determine the probability that a property of a sample of data is true for the whole population.
k-fold cross-validation	Validation technique where data is divided into several parts ("folds"), and each part is used to validate a model fit to the remaining parts. Often a more robust validation approach than splitting data into training and validation sets.
Likelihood	Probability that a model with specific parameter values would generate the actual outcomes in the data.
Maximum likelihood	A method that finds the set of parameter values for which a model is most likely to generate the actual values of the data.
Missing data	Values of data that are missing from a data set
Random effects	Patterns that appear to occur in a subset of data, but only exist due to

	random variability in the data and are not part of the system. (Note that there is a different statistical definition for this phrase too.)
Real effects	Actual patterns in the system being modeled. Ideally, good models will reveal real effects.
Sum-of-squared errors	Sum of the squares of all the differences between data and model output. In regression, this is a measure of variance.
Test data/test set	Portion of the data used to assess the effectiveness of a model once built.
Training data/training set	Portion of the data to build/fit a model. Normally, most of the data is used for training.
Validation	Measuring a model's effectiveness on data that was not used to build/train/fit the model. If there is a large difference between a model's effectiveness on a validation set of data and its effectiveness on the training set of data, it is evidence that the model may be overfit.
Validation data/validation set	Portion of the data used for validation of a model and compare between models.

NON-PARAMETRIC TESTS	***LESSON 16.1***
Mann-Whitney test	Nonparametric test to determine whether medians of two independent or unpaired samples (possibly of different size) are the same. Also called Wilcoxon sum rank test.
McNemar's test	Nonparametric test for comparing paired samples where the output is yes/no (or A/B, or 0/1, etc.).
Nonparametric test	Statistical test that makes no assumptions about the population distribution from which the data is sampled. Nonparametric tests often focus on the median.
Paired samples	Data with two different outcomes for each data point. Often helpful for comparing the method that generated outcome #1 with the method that generated outcome #2 to see which is better.
Parametric test	Statistical test that assumes the data being tested is sampled from a distribution governed by certain parameter(s). Parametric tests often focus on the mean.
Wilcoxon signed rank test (one sample)	Nonparametric test for a single response, to determining whether the median is different from a specific value.
Wilcoxon signed rank test	Nonparametric test for comparing the medians of paired samples

(paired samples)	where the output is quantitative.
OPTIMIZATION	***LESSONS 15.1-15.8, 16.3***
Approximate dynamic program	Dynamic programming model where the value functions are approximated.
Arc	Connection between two nodes/vertices in a network. In a network model, there is a variable for each arc, equal to the amount of flow on the arc, and (optionally) a capacity constraint on the arc's flow. Also called an edge.
Assignment problem	Network optimization model with two sets of nodes, that finds the best way to assign each node in one set to each node in the other set.
Bellman's equation	Equation used in dynamic programming that ensures optimality of a solution.
Binary integer program	Integer program where all variables are binary variables.
Binary variable	Variable that can take just two values: 0 and 1.
Chance constraint	A probability-based constraint. For example, a standard linear constraint might be $Ax \leq b$. A similar chance constraint might be $\Pr(Ax \leq b) \geq 0.95$.
Clique	A set of nodes where each pair is connected by an arc.
Concave function	A function f() where for every two points x and y, $f(cx + (1-c)y) \geq f(cx + (1-c)y)$ for all c between 0 and 1. In two dimensions, this means if the points $(x, f(x))$ and $(y, f(y))$ are connected with a straight line, the line is always below [or equal to] the function's curve between those two points. If $f()$ is concave, then $-f()$ is convex.
Constant	A number that remains the same.
Constraint	Part of an optimization model that describes a restriction on the solution (the values of the variables).
Convex function	A function f() where for every two points x and y, $f(cx + (1-c)y) \leq f(cx + (1-c)y)$ for all c between 0 and 1. In two dimensions, this means if the points $(x, f(x))$ and $(y, f(y))$ are connected with a straight line, the line is always above [or equal to] the function's curve between those two points. If $f()$ is convex, then $-f()$ is concave.
Convex optimization model	An optimization model where the objective function is to minimize a convex function (or maximize a concave function) and the constraints

	define a convex set of feasible solutions.
Convex quadratic function	A second-order polynomial function that is convex.
Convex quadratic program	A mathematical program where a convex quadratic function of the variables is minimized, subject to linear constraints.
Convex set	A set of points for which a straight line drawn between any two points in the set, stays inside the set. A circle is a convex set. A set shaped like the letter "U" is not convex; the line between the two points on top goes outside of the set.
Decision	Choice of action.
Diet problem	Classical optimization model for finding the least-costly set of foods that meets all dietary requirements.
Dynamic programming	Optimization approach that involves making a sequence of decisions over time, based on the current state of a system.
Edge	Connection between two nodes/vertices in a network. In a network model, there is a variable for each edge, equal to the amount of flow on the arc, and (optionally) a capacity constraint on the edge's flow. Also called an arc.
Feasible solution	A solution that satisfies a set of constraints.
Fixed charge	In optimization models, a cost that depends only on whether something happens, but not how much – for example, a transaction cost for buying or selling stock that is the same regardless of how many shares are bought or sold.
Flow	In a network model, the amount sent from one node to another along an arc. In network models, there is a variable for each arc, equal to the amount of flow on the arc, and (optionally) a capacity constraint on the arc's flow.
Global optimum/maximum/minimum	A solution that achieves the best objective value among all of the feasible solutions; sometimes also used to refer to the best objective value achievable among a set of feasible solutions.
Graph	Among other definitions, another name for a network.
Greedy algorithm	Algorithm that makes the immediately-best choice at each step.
Improving direction	Vector of changes to a solution to an optimization problem, such that the objective function gets better when moving the solution some distance in the vector's direction.
Initialization	Setting starting values in an algorithm, or setting the first solution value for an "direction/step-size" optimization algorithm.

Integer program	Optimization model where the objective function is a linear function of the variables, the constraints are linear equations and/or linear inequalities in terms of the variables, and some or all variables are restricted to have integer values.
Iterate	Repeat the same steps of a process.
Linear equation	Equation where a linear function is set equal to a constant or another linear function.
Linear function	Weighted sum of variables, plus a constant: $a_0 + \sum_{i=1}^{m} a_i x_i$.
Linear inequality	Inequality where a linear function is set to be greater-than-or-equal-to or less-than-or-equal-to a constant or another linear function.
Linear program	An mathematical programming model where the objective function is a linear function of the variables, and the constraints are linear equations and/or linear inequalities in terms of the variables.
Local optimum/maximum/minimum	A solution that achieves a better objective value than any feasible solutions that are close to it; sometimes also used to refer to that solution's objective value.
Louvain algorithm	Algorithm for finding highly-connected communities in networks.
Markov decision process	Markov chain model where decisions are made at some states, and state transitions have associated rewards.
Mathematical programming	Mathematical optimization, often using variables, constraints, and objective function.
Maximization problem	Optimization model where the objective is to find the feasible solution that maximizes the value of the objective function.
Maximum flow problem	Network optimization model that finds the most flow that can be sent from one specific node to another.
Minimization problem	Optimization model where the objective is to find the feasible solution that minimizes the value of the objective function.
Modularity	Measure of the density of connections between communicates in a network.
Most optimal	Please don't say this (or "more optimal"). "Optimal" means "best", and "most best" or "more best" are not proper English.
Network	Model where locations (nodes or vertices) are connected by arcs or edges, with flow on the arcs from node to node.
Network optimization problem	Optimization problem that can be modeled as a network with nodes and arcs, where each variable represents the flow on an arc, with constraints to ensure that the flow into each node equals the flow out

	of it, and to put a capacity on the flow on each arc.
Node	Location in a network. In a network model, there is a constraint for each node to ensure that the incoming flow equals the outgoing flow. Also called a vertex.
Non-convex program	Optimization model where the constraint set is not convex, and/or the objective function is to minimize a nonconvex function or to maximize a nonconcave function.
Non-negativity constraints	Constraints that require variables to be greater than or equal to zero.
Objective function	Part of an optimization model that measures the quality of a solution (the values of the variables).
Optimal	Best possible, while satisfying all constraints.
Optimal solution	A solution that satisfies a set of constraints, and has the best-possible objective value.
Optimization	Finding the values of variables/decisions that yield the best value of an objective function while satisfying a set of constraints (restrictions).
Robust solution	A solution that whose worst-case outcome over all possible scenarios is least bad.
Scenario	Specific case/instance of an uncertain outcome; one approach to stochastic optimization is to optimize over a number of scenarios simultaneously.
Shortest path problem	Network optimization model that finds the shortest route in a network from one specific node to another.
Solution (in the optimization sense)	A vector of values, one for each variable in an optimization model.
State	Description of a system's condition.
Step size	Distance to move in an improving direction, to get to a new solution given a current solution and an improving direction. The new solution is equal to the old solution, plus the product of the improving direction and step size.
Stochastic dynamic program	Dynamic program where the outcome of one or more decisions is determined according to probabilities.
Stochastic optimization	An optimization model that accounts for randomness or uncertainty.
Uncertainty	Lack of knowledge about a data value, parameter value, outcome, etc.
Variable (optimization sense)	A decision that an optimization model suggests a value for.
Variable (statistics sense)	An attribute whose value can differ for different data points.

Vertex	Location in a network. In a network model, there is a constraint for each vertex to ensure that the incoming flow equals the outgoing flow. Also called a node.

PROBABILITY-BASED MODELS	***LESSONS 13.5-13.8, 16.2***
Action	In ARENA, something that is done to an entity.
Arrival rate	Expected number of arrivals of people, things, etc. per unit time -- for example, the expected number of truck deliveries per hour to a warehouse.
Balking	An entity arrives to the queue, sees the size of the line (or some other attribute), and decides to leave the system.
Bayes' theorem/Bayes' rule	Fundamental rule of conditional probability: $P(A\|B) = \frac{P(B\|A)P(A)}{P(B)}$.
Continuous-time simulation	A simulation that models a system continuously, at every instant of time; continuous-time simulation models are often based on differential equations.
Decision point	Place in a simulation where there is a branch (or decision to be made or observed).
Deterministic simulation	Simulation with no randomness/uncertainty, so results are the same each run.
Discrete-event simulation	A simulation that models a system that changes when specific events occur.
Empirical Bayes model	Model that uses Bayes' theorem to update an initial guess/distribution based on observed data.
Entity	A person/thing moving through a simulation.
FIFO	First-in, first-out: The first entity to join a queue is the first one to come out -- for example, a supermarket checkout line.
Interarrival time	The time between two consecutive arrivals of people, things, etc. -- for example, the time between consecutive phone calls to a service hotline.
Kendall notation	Notation to describe various types of queuing models -- for example, M/M/c (a queue with Poisson arrivals, exponentially-distributed service times, and c identical servers).
LIFO	Last-in, first-out: The last entity to join a queue is the first one to come out -- for example, a stack of papers.
Markov chain	Process where a system changes its state in a way that depends only

	on its current state.
Memoryless (Markov chain)	Property that the next state of the system is dependent only on the current state, not any previous states.
Module	In ARENA, a building-block of a simulation, or the process, resource, etc. it represents.
Queue	A line of people, things, etc. waiting to go through or be processed/served by a resource -- for example, an airport security line.
Queuing	The mathematical study of queues.
Replication	Running a stochastic simulation multiple times to sample the distribution of possible simulation results. "A replication" also refers to a single one of many runs of the simulation.
Resource	In ARENA, the "doers" – for example, a call center worker at a queue.
Service rate	Rate at which entities are processed.
Simulation	A model that imitates the operation or behavior of a real system.
Steady state	In a Markov chain, having the same probability distribution of being in each state, before and after a transition.
Stochastic simulation	Simulation that includes randomness/uncertainty, so results can be different each run.
Transition matrix	Matrix of transition probabilities.
Transition probability	Probability of moving from current state i to next state j, often denoted p_{ij}.
Validation (of simulation)	Making sure that simulation results are similar-enough to those of the real system being simulated, so the simulation can be used to analyze the real system.

PROBABILITY DISTRIBUTIONS	***LESSONS 13.1-13.4***
Bernoulli distribution	Discrete probability distribution where the outcome is binary, either 0 or 1. Often, 1 represents success and 0 represents failure. The probability of the outcome being 1 is p and the probability of outcome being 0 is $q = 1 - p$, where p is between 0 and 1.
Bias	Systematic difference between a true parameter of a population and its estimate.
Binomial distribution	Discrete probability distribution for the exact number of successes, k, out of a total of n iid Bernoulli trials, each with probability p:

	$\Pr(k) = \binom{n}{k} p^k (1-p)^{n-k}$.
Distribution-fitting	Determining whether a set of data seems to follow a certain probability distribution, or determining which of several distributions the data is close to.
Exponential distribution	A continuous probability distribution of the time between events: $f(x) = \lambda e^{-\lambda x}$. If the number of events in a fixed time follows the Poission distribution, then the time between them has the exponential distribution. The exponential distribution has the memoryless property.
Geometric distribution	Discrete probability distribution of the number of iid Bernoulli trials, each with success probability p, before the first success: $\Pr(k) = (1-p)^k p$. Also can be defined as the total number of trials through the first success (so $\Pr(k) = (1-p)^{k-1} p$). To find the number of trials before the first failure, a similar distribution would be $\Pr(k) = p^k (1-p)$.
iid	Independent and identically distributed.
Independent	A is "independent" of B if the probability or probability distribution of A is not affected by B. For example, whether a coin flip is heads or tails is (I assume) independent of the number of fish in the ocean exactly 100 years ago to this day, but the temperature today is not independent of the temperature yesterday (if it was hot yesterday, it's more likely to be hot today too, etc.).
Independent and identically distributed (iid)	Things that follow the same probability distribution, including the same parameter(s), and whose values are independent of each other. For example, multiple flips of the same coin are iid.
Lower tail	Lowest-value part of a distribution
Memoryless (distribution)	Probability distributions where the past history of outcomes does not influence the probability of the outcome of future events. The exponential and geometric distributions have this property.
Normal distribution	Continuous probability distribution: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Model error is often assumed to be normally distributed (for example, in linear regression).
Poisson distribution	A discrete probability distribution of the number of iid events happening within a fixed time: $\Pr(k) = \frac{\lambda^k e^{-\lambda}}{k!}$. If the time between the events follows the exponential distribution, then the number of events follows the Poisson distribution.

Q-Q plot	Quantile-quantile plot -- a plot comparing the quantiles of two data sets, or one data set and a distribution, to see whether they might have a common distribution.
Tail(s)	Highest and lowest-value parts of a distribution.
Upper tail	Highest-value part of a distribution
Weibull distribution	Continuous probability distribution that is often used to model the time until failure of a device, component, etc.: $f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(\frac{x}{\lambda})^k}$ for $x \geq 0$.

REGRESSION	***LESSONS 8.1, 8.3-8.6, 9.1, 10.1-10.4, 10.7***
Adjusted R-squared/Adjusted R^2	Variant of R^2 that encourages simpler models by penalizing the use of too many variables.
Area under curve/AUC	Area under the ROC curve; an estimate of the classification model's accuracy. Also called concordance index.
Bayesian regression	Regression model that incorporates estimates of how coefficients and error are distributed.
Box-Cox transformation	Transformation of a non-normally-distributed response to a normal distribution.
Branching	Splitting a set of data into two or more subsets, to each be analyzed separately.
CART	Classification and regression trees.
Classification tree	Tree-based method for classification. After branching to split the data, each subset is analyzed with its own classification model.
Concordance index	Area under the ROC curve; an estimate of the classification model's accuracy. Also called AUC.
Decision tree	Tree-based method for decision-making. After branching to split the data, each subset is analyzed with its own decision model (or just has its own decision applied).
Earth	Name of many implementations of multi-adaptive regression spline (MARS) model, because "MARS" is a trademark.
Elastic net	Combination of lasso and ridge regression.
Forest	A set of multiple trees. Just like in real life.
Interaction term	Variable in a model that is the combination of two or more other variables; for example, if x_1 and x_2 are variables, $(x_1 x_2)$ is an

	interaction term/interaction variable.
k-Nearest-Neighbor regression	Regression model where a data point's response is estimated based on the responses of the k nearest data points with known response.
Knot	Point where pieces of a spline regression meet.
Lasso/Lasso regression	Method for limiting the number of variables in a model by limiting the sum of all coefficients' absolute values. Can be very helpful when number of data points is less than number of factors.
Leaf	In a tree model, a subset of data from which there is no branching.
Linear regression	Regression model where the relationships between attributes and a response are modeled as linear functions: $y = a_0 + \sum_{i=1}^{m} a_i x_i$.
Logistic regression	Regression model that uses an exponential function of variables to estimate a response that is either between 0 and 1, or must be equal to 0 or 1: $y = \frac{1}{1+e^{-\left(a_0+\sum_{i=1}^{m} a_i x_i\right)}}$. Also called a logit model.
Logit model	Regression model that uses an exponential function of variables to estimate a response between 0 and 1: $y = \frac{1}{1+e^{-\left(a_0+\sum_{i=1}^{m} a_i x_i\right)}}$. Also called a logistic regression.
MARS	Multi-adaptive regression splines.
Multi-adaptive regression splines (MARS)	Specific regression spline model that has become commonly-used. Abbreviation "MARS" is a trademark, so many versions are called "earth".
p-value	(1) In hypothesis testing, probability that results at least as extreme as those in the data would be observed if the null hypothesis is true. (2) In regression, probability that results at least as extreme as those in the data would be observed if the coefficient of a variable is zero.
p-value fishing	Testing many different hypotheses hoping to find one with a low p-value. This is a bad practice; if enough things are tested, it's likely one will have a low p-value due to randomness, but that doesn't mean it's a real effect.
Poisson regression	Regression that assumes the response has a Poisson distribution.
Pruning	Removing a branch from a tree.
Pseudo-R-squared/Pseudo-R^2	Measure similar to R^2 used for nonlinear regression models where R^2 cannot be calculated.
R-squared/R^2	Measure of linear regression model quality, the fraction of variance in the response that is explained by the model. Also called coefficient of determination.

Random forest	Machine learning model that creates many different trees and returns their mean output. Can be used with classification trees, regression trees, decision trees.
Receiver operating characteristic curve (ROC curve)	Graph that plots the true positive rate against the false positive rates for different classification cutoff thresholds.
Regression	Statistical model that describes relationships between variables, and/or predicts future values of a response..
Regression splines	Regression model where different functions are used for different ranges of the data. Also called spline regression.
Regression tree	Tree-based method for regression. After branching to split the data, each subset is analyzed with its own regression model.
Ridge regression	Method of regularization by limiting the sum of the squares of the coefficients. Will reduce the magnitude of coefficients, not the number of variables chosen.
ROC curve	Receiver operating characteristic curve.
Root	The first, complete data set in a tree model.
Spline regression	Regression model where different functions are used for different ranges of the data. Also called regression splines.
Transformation	A mapping of points from one space to another.
Tree	Iterative split (branching) of a data set into more-specific subsets that each are modeled separately. Often used for classification, regression, and decision-making. Also can be used to solve optimization problems.

TIME SERIES MODELS	***LESSONS 7.1-7.6***
Additive seasonality	Seasonal effect that is added to a baseline value (for example, "the temperature in June is 10 degrees above the annual baseline").
ARIMA	Autoregressive integrated moving average.
Autoregression	Regression technique using past values of time series data as predictors of future values.
Autoregressive integrated moving average (ARIMA)	Time series model that uses differences between observations when data is nonstationary. Also called Box-Jenkins.
Differencing	Using the difference of successive values in time series data, rather than the values themselves. Sometimes nonstationary data will have stationary differences.

Double exponential smoothing	Two-parameter exponential smoothing technique that incorporates trend.
Exponential smoothing	Data smoothing technique in which older observations are assigned exponentially decresing weights, so more emphasis is given to recent observations.
GARCH	Generalized autoregressive conditional heteroscedasticity.
Generalized autoregressive conditional heteroscedasticity (GARCH)	Autoregressive method used to model variance in time series data.
Holt-Winters method	Three-parameter exponential smoothing technique that incorporates trend and seasonality; also called triple exponential smoothing. Also called Winters' method.
Moving average	Smoothing technique that replaces data values with the mean of a number of consecutive observed values.
Multiplicative seasonality	Seasonal effect that is multiplied by a baseline value (for example, "the temperature in June is 20% higher than the annual baseline").
Seasonality/cycles	Repeating pattern in data values over time, often at consistent intervals (for example, temperature variations throughout the year that repeat each year at about the same time).
Seasonality length/cycle length	Fixed time period at which cycles/seasonalities repeat themselves.
Single exponential smoothing	Exponential smoothing technique with just one parameter, that does not incorporate trend or seasonality.
Smoothing	Time series analysis technique to help filter out underlying randomness/noise. Examples include moving average, exponential smoothing, and ARIMA.
Smoothing constant	Parameter in exponential smoothing to determine the relative importance of recent observations and previous estimates. Smoothing constants are between 0 and 1; a higher value indicates more reliance on observation, and a lower value indicates more reliance on previous estimates.
Stationary process	Process whose joint probability distribution and statistical properties (mean, variance, autocorrelation, etc.) do not vary with time. Examples include data with trends or cycles.
Trend	Increase or decrease in data values over time.
Triple exponential smoothing	Three-parameter exponential smoothing technique that incorporates trend and seasonality; also called Winters' method or Holt-Winters.

Winters' method	Three-parameter exponential smoothing technique that incorporates trend and seasonality; also called triple exponential smoothing. Also called Holt-Winters.

VARIABLE SELECTION	***LESSONS 11.1-11.3***
Backward elimination	Variable selection process that starts with all variables and then iteratively removes the least-immediately-relevant variables from the model.
Elastic net	Combination of lasso and ridge regression.
Forward selection	Variable selection process that starts with no variables and then iteratively adds the most-immediately-relevant variables to the model.
Lasso/Lasso regression	Method for limiting the number of variables in a model by limiting the sum of all coefficients' absolute values. Can be very helpful when number of data points is less than number of factors.
Overfitting	Building a model that describes random effects instead of or in significant addition to the real effects; often caused by having too many factors or parameters compared to the number of data points. Overfitted models will have high prediction errors.
Regularization	Addition of term(s) to the model to reduce model complexity or overfitting. For example, adding a penalty to the objective function in regression can help reduce overfitting (see ridge regression).
Ridge regression	Method of regularization by limiting the sum of the squares of the coefficients. Will reduce the magnitude of coefficients, not the number of variables chosen.
Simplicity (of a model)	Having fewer parameters; opposite of complexity of a model. Often helpful for avoiding overfitting and increasing interpretability.
Stepwise regression	Variable selection process that can combine forward selection and backward regression.
Variable selection	Process of selecting the best subset of predictors to explain variance in data; involves eliminating unnecessary or redundant or less-important variables from a potential set of predictors.

OTHER TOPICS	***LESSONS 1.1, 4.2, and OTHER ASSORTED LESSONS***
1-norm	Similar to rectilinear distance; measures the sum of the lengths of each dimension of a vector from the origin. If $z = (z_1, z_2, \dots, z_m)$ is a vector in an m-dimensional space, then its 1-norm is

	$\sqrt[1]{\lvert z_1\rvert^1 + \lvert z_2\rvert^1 + \cdots + \lvert z_m\rvert^1} = \lvert z_1\rvert + \lvert z_2\rvert + \cdots + \lvert z\rvert = \sum_{i=1}^{m} \lvert z_i\rvert$.
2-norm	Similar to Euclidian distance; measures the straight-line length of a vector from the origin. If $z = (z_1, z_2, \ldots, z_m)$ is a vector in an m-dimensional space, then its 2-norm is $\sqrt[2]{(z_1)^2 + (z_2)^2 + \cdots + (z_m)^2} = \sqrt[2]{\sum_{i=1}^{m} (z_i)^2}$.
Convex hull (of a set of points)	Smallest convex shape that the set of points is contained in.
Descriptive analytics	Loosely speaking, the use of analytics to explain or describe what has happened.
Distance	How far it is between two points -- but there are different ways to measure it (see Minkowski distance).
Elbow diagram	A graph of improvement in function value as something else (e.g., number of clusters) increases or decreases; the spot where improvement levels out is the "elbow".
Error (per data point)	The difference (or absolute difference, squared difference, or other measure) between the estimate of a piece of data and its true value.
Error (total over data set)	The total of all errors in a data set.
Euclidian distance/straight-line distance	The length of a straight line (the 2-norm distance) between two points. If $x = (x_1, x_2, \ldots, x_m)$ and $y = (y_1, y_2, \ldots, y_m)$ are two points in an m-dimensional space, then the Euclidian distance between them is $\sqrt[2]{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_m - y_m)^2} = \sqrt[2]{\sum_{i=1}^{m} (x_i - y_i)^2}$.
Fitting	Finding a model (including, if appropriate, a probability distribution) that is a good description of real effects in a set of data. The model is sometimes called a “fit”.
Heteroscedasticity	When the variability of a response is different across the range of predictor values.
Infinity-norm	Specific case of p-norm when $p = \infty$. Sounds weird, but it just reduces to the largest of the dimensions. If $z = (z_1, z_2, \ldots, z_m)$ is a vector in an m-dimensional space, then its ∞-norm is $\max_i \lvert z_i\rvert$. If $x = (x_1, x_2, \ldots, x_m)$ and $y = (y_1, y_2, \ldots, y_m)$ are two points in an m-dimensional space, then the ∞-norm distance between them is $\max_i \lvert x_i - y_i\rvert$.
Linear combination	The weighted sum of things. For example, if $x_1, x_2, \ldots, x_m$ are factors, then $a_1x_1 + a_2x_2 + \cdots + a_mx_m$ is a weighted sum of them for any numbers $a_1, a_2, \ldots, a_m$.
Manhattan distance	The sum of the lengths in each dimension between two points (the 1-norm distance). If $x = (x_1, x_2, \ldots, x_m)$ and $y = (y_1, y_2, \ldots, y_m)$ are two points in an m-dimensional space, then the rectilinear distance

	between them is $\sqrt[1]{\lvert x_1 - y_1\rvert^1 + \lvert x_2 - y_2\rvert^1 + \cdots + \lvert x_m - y_m\rvert^1}$ = $\lvert x_1 - y_1\rvert + \lvert x_2 - y_2\rvert + \cdots + \lvert x_m - y_m\rvert = \sum_{i=1}^{m}\lvert x_i - y_i\rvert$. Also called Rectilinear or 1-norm distance.
Minkowski distance (of order p)	The p-norm distance between two points. If $x = (x_1, x_2, \ldots, x_m)$ and $y = (y_1, y_2, \ldots, y_m)$ are two points in an m-dimensional space, then the Minkowski distance of order p between them is $\sqrt[p]{\lvert x_1 - y_1\rvert^p + \lvert x_2 - y_2\rvert^p + \cdots + \lvert x_m - y_m\rvert^p} = \sqrt[p]{\sum_{i=1}^{m}\lvert x_i - y_i\rvert^p}$.
Model (mathematical)	A mathematical description of a system. Because real-life systems are complex, mathematical models of them are only approximate. In analytics, the term "model" is used in at least three different ways: (1) A general type of mathematical approach, like "regression"; (2) A general type of mathematical approach with specific parameters, like "regression using credit score and income as predictors"; (3) A general type of mathematical approach with specific parameters and values for the parameters, like "regression, with the prediction equal to 100,000, plus 100 times credit score, plus 3 times income".
Multiplier	A term that something is multiplied by. For example, to change units from meters to centimeters, the multiplier is 100.
Norm/distance norm	A function that measures the size/length of a vector and satisfies some basic technical properties that are beyond the scope of this course. In this course, we focus on Minkowski norm (or p-norm).
Order of magnitude	The relative size of something, often denoted by multiples of 10 so that difference in the order of magnitude of two numbers is the difference in how many digits they haves. So, loosely speaking, a 2-digit number is one order of magnitude smaller than a 3-digit number, a 7-digit number is two orders of magnitude smaller than a 9-digit number, two 4-digit numbers have the same order of magnitude, etc.
Orthogonal	At right angles to one another (like "perpendicular" but generalized to more dimensions). Statistically, if two attributes are orthogonal then they are independent.
Outlier	A data point or set of points that's far from the rest in one way or another (see point outlier, contextual outlier, collective outlier).
Overfitting	Building a model that describes random effects instead of or in significant addition to the real effects; often caused by having too many factors or parameters compared to the number of data points. Overfitted models will have high prediction errors.
p-norm	Measures vector length similar to the Minkowski distance of order p. If $z = (z_1, z_2, \ldots, z_m)$ is a vector in an m-dimensional space, then its p-norm is $\sqrt[p]{\lvert z_1\rvert^p + \lvert z_2\rvert^p + \cdots + \lvert z_m\rvert^p} = \sqrt[p]{\sum_{i=1}^{m}\lvert z_i\rvert^p}$.

Parameter	A constant whose value determines something about a system, expression, etc. For example, if we remove a variable from a regression model whenever its p-value is "too high", above P, then P is a parameter, and setting it to different values can mean we get different models.
Perturbation	A change (usually small) from the actual or expected value of something.
Prediction	Estimate of what will happen in the future, or of something unknown (e.g., missing data) that happened.
Predictive analytics	Loosely speaking, the use of analytics to estimate or predict what will happen.
Prescriptive analytics	Loosely speaking, the use of analytics to suggest or prescribe what's best to do.
Rectilinear distance	The sum of the lengths in each dimension between two points (the 1-norm distance). If $x = (x_1, x_2, \dots, x_m)$ and $y = (y_1, y_2, \dots, y_m)$ are two points in an m-dimensional space, then the rectilinear distance between them is $\sqrt[1]{\lvert x_1 - y_1\rvert^1 + \lvert x_2 - y_2\rvert^1 + \cdots + \lvert x_m - y_m\rvert^1}$ = $\lvert x_1 - y_1\rvert + \lvert x_2 - y_2\rvert + \cdots + \lvert x_m - y_m\rvert = \sum_{i=1}^{m} \lvert x_i - y_i\rvert$. Also called Manhattan or 1-norm distance.
Threshold	A value that denotes the difference between something happening or not happening. For example, if "whenever p is greater than 0.15, we include the corresponding variable in a regression model" then 0.15 would be the "threshold" value of p that differentiates between including and not including the variable.
Transformation	A mapping of points from one space to another.