Translate_Stats_ML Cheat Sheet

by Pranay Kanwar (r4um) via cheatography.com/57680/cs/18133/

Terminology

Statistics	Machine learning	Notes
data point, record, row of data	example, instance	Both domains also use "observation," which can refer to a single measurement or an entire vector of attributes depending on context.
response variable, dependent variable	label, output	Both domains also use "target." Since practically *all* variables depend on other variables, the term "dependent variable" is potentially misleading.
variable, covariate, predictor, independent variable	feature, side information, input	The term "independent variable" exists for historical reasons but is usually misleading--such a variable typically depends on other variables in the model.
regressions	supervised learners, machines	Both estimate output(s) in terms of input(s).
estimation	learning	Both translate data into quantitative claims, becoming more accurate as the supply of relevant data increases.
hypothesis ≠ classifier	hypothesis	In both statistics and ML, a hypothesis is a scientific statement to be scrutinized, such as "The true value of this parameter is zero." In ML (but not in statistics), a hypothesis can also refer to the prediction rule that is output by a classifier algorithm.
bias ≠ regression intercept	bias	Statistics distinguishes between (a) bias as form of estimation error and (b) the default prediction of a linear model in the special case where all inputs are 0. ML sometimes uses "bias" to refer to both of these concepts, although the best ML researchers certainly understand the difference.

Terminology (cont)

Maximize the likelihood to estimate model parameters	If your target distribution is discrete (such as in logistic regression), minimize the entropy to derive the best parameters. If your target distribution is continuous, fine, just maximize the likelihood.	For discrete distributions, maximizing the likelihood is equivalent to minimizing the entropy.
Apply Occam's razor, or encode missing prior information with suitably uninformative priors	Apply the principle of maximum entropy.	The principle of maximum entropy is conceptual and does not refer to maximizing a concrete objective function. The principle is that models should be conservative in the sense that they be no more confident in the predictions than is thoroughly justified by the data. In practice this works out as deriving an estimation procedure in terms of a bare-minimum set of criteria as exemplified here or here.
logistic/multinomial regression	maximum entropy, MaxEnt	They are equivalent except in special multinomial settings like ordinal logistic regression. Note that maximum entropy here refers to the principle of maximum entropy, not the form of the objective function. Indeed, in MaxEnt, you minimize rather than maximize the entropy expression.
X causes Y if surgical (or randomized controlled) manipulations in X are correlated with changes in Y	X causes Y if it doesn't obviously not cause Y. For example, X causes Y if X precedes Y in time (or is at least contemporaneous)	The stats definition is more aligned with common-sense intuition than the ML one proposed here. In fairness, not all ML practitioners are so abusive of causation terminology, and some of the blame belongs with even earlier abuses such as Granger causality.

Terminology (cont)

structural equations model	Bayesian network	These are nearly equivalent mathematically, although interpretations differ by use case,as discussed.
sequential experimental design	active learning, reinforcement learning, hyperparameter optimization	Although these four subfields are very different from each other in terms of their standard use cases, they all address problems of optimization via a sequence of queries/experiments.

Source https://insights.sei.cmu.edu/sei_blog/2018/11/translating-between-statistics-and-machine-learning.html

By **Pranay Kanwar** (r4um)
cheatography.com/r4um/