Probability and Statistics for Computer Vision

This note clarify how statistics is essential of the computer vision and what is the most subjects from statistics is used directly in the computer vision.

Firstly, probability rules are almost used which is the measure of the likelihood that an event will occur represented in numbers between 0 to 1. Because it depends on the existence of the events or the likelihood of events occurs and how this events effects on the cases probability of the code story. and this is just introducing the random variable which represents the outcome of our interest.

Probability density function: is a function whose value at any given sample which is set of possible values taken by the random variable. So the density of a continuous random variable represented as a function where area under the curve shows the probability of the RV itself.

Joint probability: the probability of the two or more event with each other despite the separated history probability of each one. That is why they calculates the likelihood of two (or more!) events occurring together and at the same point in time.

$$Pr(x) = \int Pr(x, y)\, dy \qquad \text{(Continuous)}$$

$$Pr(x) = \sum_{y} Pr(x, y) \qquad \text{(Discrete)}$$

Marginalization: Which is trying to think about all the possible situations for the state you are interested in. So, it is the way to ignore a certain random variable to get a probability with less random variables.

$$Pr(x \mid y = y^*) = \frac{Pr(x, y = y^*)}{\int Pr(x, y = y^*)\, dx} = \frac{Pr(x, y = y^*)}{Pr(y = y^*)}$$

← Normalization Term (By dividing with it's own area, normalizing to 1.)

Conditional Probability: The probability of an event given that (by assumption, presumption, assertion or evidence) another event has occurred. restricting the situation given a certain condition.

Given y, probability of x

Joint probability of x & y

$$Pr(x \mid y) = \frac{Pr(x, y)}{Pr(y)}$$

Probability of y (Normalization Term)

Bayes' Rule: The rule that ties posterior probability with likelihood, prior, and evidence.

Why we use statistics in machine learning?

1. For understanding and benchmarking data as well as for model validation.
2. Many important algorithms build by the statistics theory such as logistic regression.
3. provide solutions to entirely new class of problems (such as computer vision) in deep learning.

The difference between the machine learning and the statistics:

Both Statistics and ML are often utilized together to provide a robust solution.

	Statistics	Machine Learning
Approach	Data Generating Process	Algorithmic Model
Driver	Math, Theory	Fitting Data
Focus	Hypothesis Testing, Interpretability	Predictive Accuracy
Data Size	Any Reasonable Set	Big Data
Dimensions	Used Mostly for Low Dimensions	High Dimensional Data
Inference	Parameter Estimation, Predictions, Estimating Error Bars	Prediction
Model Choice	Parameter Significance, In-sample Goodness of Fit	Cross-validation of Predictive Accuracy on Partitions of Data
Popular Tools	R	Python
Interpretability	High	Low

@thebabar

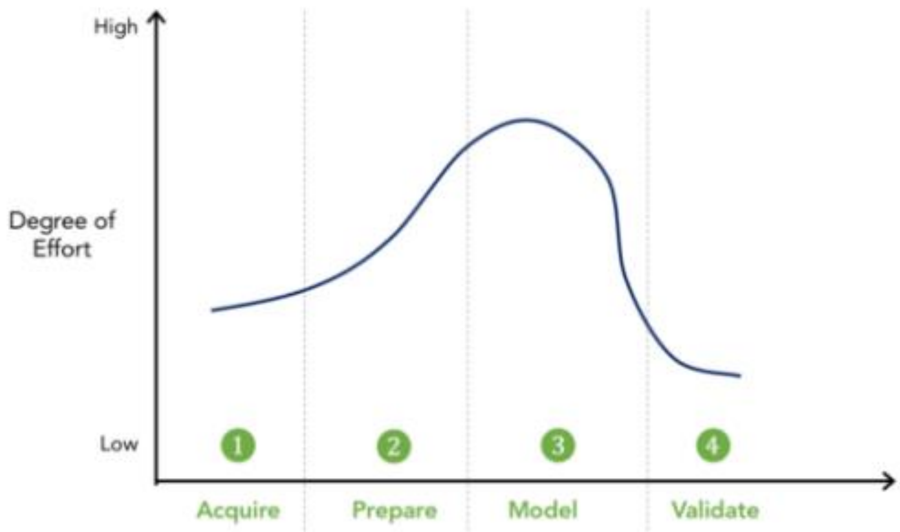
1. ML algorithms are typically far more complex than their statistical counterparts.
2. ML require design decisions to be made before the training process begins.
3. ML has complexity leads to an added layer of challenge of interpretation.

<u>Feature Engineering from the Machine Learning point of view:</u>

1. Large number of inputs
2. Unstructured data (e.g. speech) needs feature engineering as a pre-processing step before training can start
3. AutoML can generate large number of complex features to test many transformations of the data — this can add unnecessary complexity.

<u>Hyperparameters</u> are also a challenge in Machine Learning. The depth of trees in a random-forest model or the number of layers in a deep neural network, must be defined *before the training process can begin*. Decisions regarding hyperparameters, are often more complex than analogous decisions in statistical modeling.

<u>Statistics in the ML Workflow:</u>



Four Phases of ML and Statistics

Validation:

How do we validate that we have the representative sample? The accepted approach is to check p-value of a sample to ensure its over 5% confidence level (i.e. it passes the significance test). *N*umerical variables must be checked with the Kolmogorov-Smirnov test, while categorical variables need Pearson's chi-square test. For instance, Bootstrap is a resampling method.

Many times statistical modeling and ML use very similar approaches and therefore overlap with each other.

Logistic Regression — Logistic regression is one technique borrowed by machine learning from the field of statistics. It is a widely used method for binary classification problems but can also be extended to multi-class problems.

Here's a list of terms that are similar in meaning but called with different names.

Statistics	**Machine Learning**
Data Point	Instance
Covariate	Feature
Parameters	Weights
Estimation / Fitting	Learning
Regression / Classification	Supervised Learning
Clustering / Density Estimation	Unsupervised Learning
Response	Label
Test set performance	Generalization

Statistics and ML — Terms with Different Names, Similar Meaning