Learning Lounge:

Using Predictive Analytics to Solve Business Problems

Commitment Beyond Numbers

Linda Brobeck, FCAS, MAAA, CSPA

June 26, 2018

Machine Learning Overview

Machine Learning

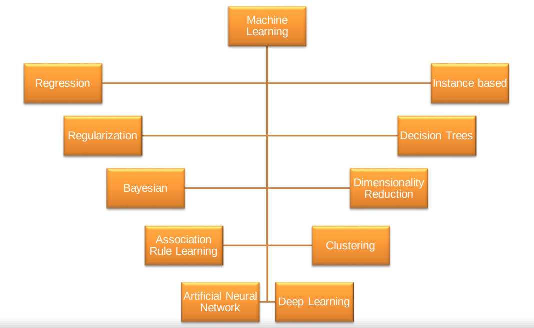
- Supervised
  - Predictive
  - Target Variable
  - Task Driven
  - Regression, Classification
- Unsupervised
  - Descriptive
  - No Target Variable
  - Data Driven
  - Clustering, Pattern Discovery, Dimension Reduction

Reinforcement – Algorithm Learns to React

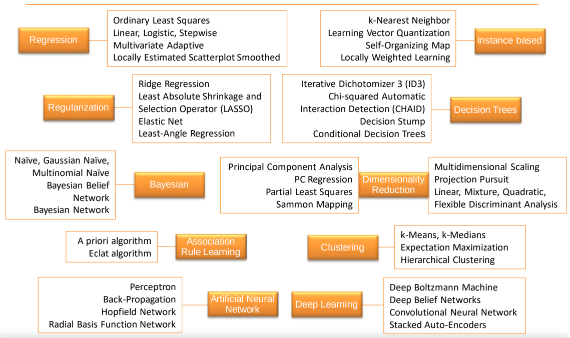
How do you define Machine Learning?

- A Analytics with reinforcement
- B Predictive modeling – any technique - no reinforcement
- C Predictive modeling beyond GLM without reinforcement
- D Other

Machine Learning Overview



Machine Learning Overview



Parametric vs Non-parametric

Parametric

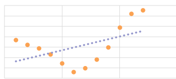
The shape of the predictor function is defined by a few parameters

- Algorithms simplify the function to a known form
- Machine learning finds coefficients

Nonparametric

The shapes of predictor functions are fully determined by the data

Generalized Linear Modeling (GLM)



- **Generalized Linear Model**
  - $\eta = X\beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$
  - $\mu = g^{-1}(\eta)$
- **Goodness-of-fit**
  - Conceptually equivalent to sum of squares in ordinary linear regression
  - Deviance: $\sum_i \frac{2wi}{\phi} \int_{\mu_i}^{y_i} \frac{y_i - \theta}{V(\theta)}\, d\theta$
- **Key Assumptions**
  - Link function g() – describes the functional relationship between the parameters and the fitted value
  - Variance function $V(\mu_i)$ – describes the relationship between the mean and variance for an observation
  - Scale Parameter ϕ - accounts for misspecification of the variance function
  - Weights w_i - the amount of weight given to an individual record

Generalized Additive Models (GAM)

Replaces estimation of linear form parameters with smooth linear or non-linear functions

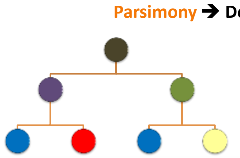
- **Generalized Additive Model**
  - $\eta = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_p(x_p)$
  - $\mu = g^{-1}(\eta)$
- **Goodness-of-fit**
  - Similar to GLMs
- **Functions f_i can be:**
  - Parametric with a specified form (i.e., a polynomial)
  - Non-Parametric
  - Each f_i can be a different function

Decision Trees - Methodology

Split data according to measures of similarity

If the Target Variable is: Categorical ➔ **Classification Tree**
Continuous ➔ **Regression Tree**

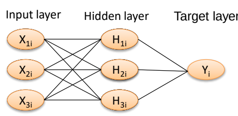
Two Competing Objectives: Purity ➔ **Measure of Variation**
Parsimony ➔ **Desire for Simple**



Neural Networks

Requires limited assumptions regarding the relationship of explanatory variables.

- Target layer regression model on a series of derived input, called hidden units
- Hidden units are regressions on the original inputs
- Regression parameters are adjusted iteratively to minimize the squared residuals



Ensembles

Combine many weak classifiers in order to strengthen the overall result

Bagging (Bootstrap Aggregating)

Many models each based on sample

Each model in the ensemble gets a vote

Boosting

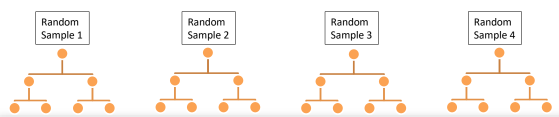
Iterative models dependent on previous model

Stacked Generalization (Blending)

Use of diverse models in combination

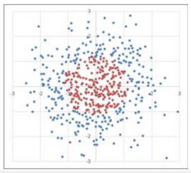
Random Forests

- Each tree is built using a random sample
  - With replacement of the observations
  - Without replacement of the explanatory variables
- The predicted target value is the mean predicted target value over the ensemble
- Perturbation (interjecting randomness) is implemented, at each node, by only searching for optimal splits among a randomly chosen subset of the explanatory variables



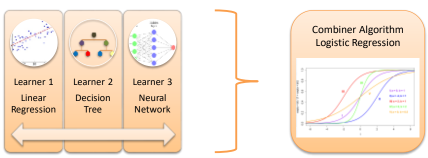
Boosting

- **Gradient Boosting**
  - Trees built sequentially, tree based on previously built because the new tree is built on the residual of prior tree(s)
- **Multiplicative Boosted Trees**
  - Multiplicative residuals
  - Multiplicative combining of trees
- **AdaBoost**
  - Adaptive boosting
  - Iteratively changes weights of training observations based on errors of previous prediction



Stacked Generalization

- Blending of many model types
- Use diverse models for the blending/stacking
- 2 stages
  1. Fit base learners to data
  2. Fit a combiner algorithm to the predictions of the base learners



Summary

- **Selecting/Combining Techniques**
  - Depends on the application/objective
  - No silver bullet
- **Machine Learning Reinforcement vs. Domain Expertise**

Questions