Introduction to Machine Learning

Ranga Raju Vatsavai, Ph.D.
Chancellors Faculty Excellence Associate Professor in Geospatial Analytics
Department of Computer Science, North Carolina State University (NCSU)

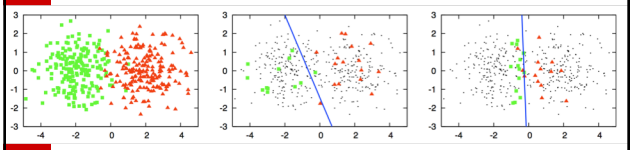
Feb. 25-27, 2019

Generative vs. Discriminative

- Generative classifiers
  - Assume some functional form for P(Y), P(X|Y)
  - Estimate parameters of P(X|Y), P(Y) directly from training data
  - Use Bayes rule to calculate P(Y|X)
  - Examples: NB, MLC/MAP, MRF, …
- Discriminative classifiers
  - Assume some functional form for P(Y|X)
  - Estimate parameters of P(Y|X) directly from training data
  - Examples: Logistics Regression, kNN, NN, …

Active Learning

- Motivation



Actual Data

Small Labeled Data
Decision Boundary is not accurate

Actively queried for additional labeled data.
Decision boundary is better

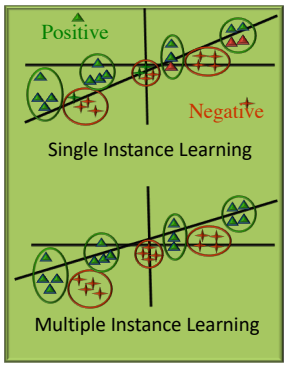
How do you query for "good" examples?

Transfer Learning

- In general, model built to learn one task do not perform on a different (or even similar) task
  - Example, take satellite image from Minnesota, build a decision tree for crop classification, and apply this decision tree to predict crop labels on North Carolina (NC) image
- Transfer learning
  - Take Minnesota (pre-trained) model and fine-tune it to work with NC crop classification

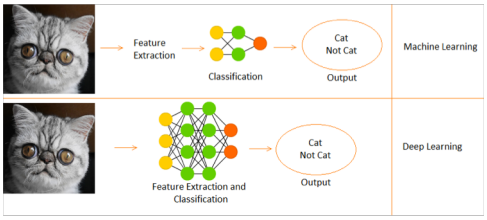
Multiple Instance Learning

- Motivation
  - Many learning tasks involves assigning a single label to "group" of instances (e.g., down-town, NCSU campus, airport, …)



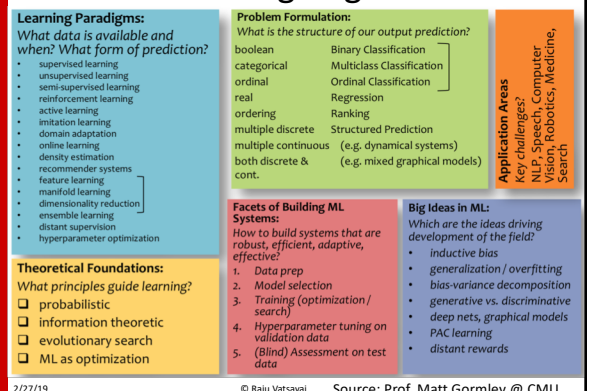
Deep Learning

- Motivation
  - Can you learn features directly from input data as "feature engineering" is costly



Source: Kaggle DL tutorial

Machine Learning: Big Picture



Source: Prof. Matt Gormley @ CMU