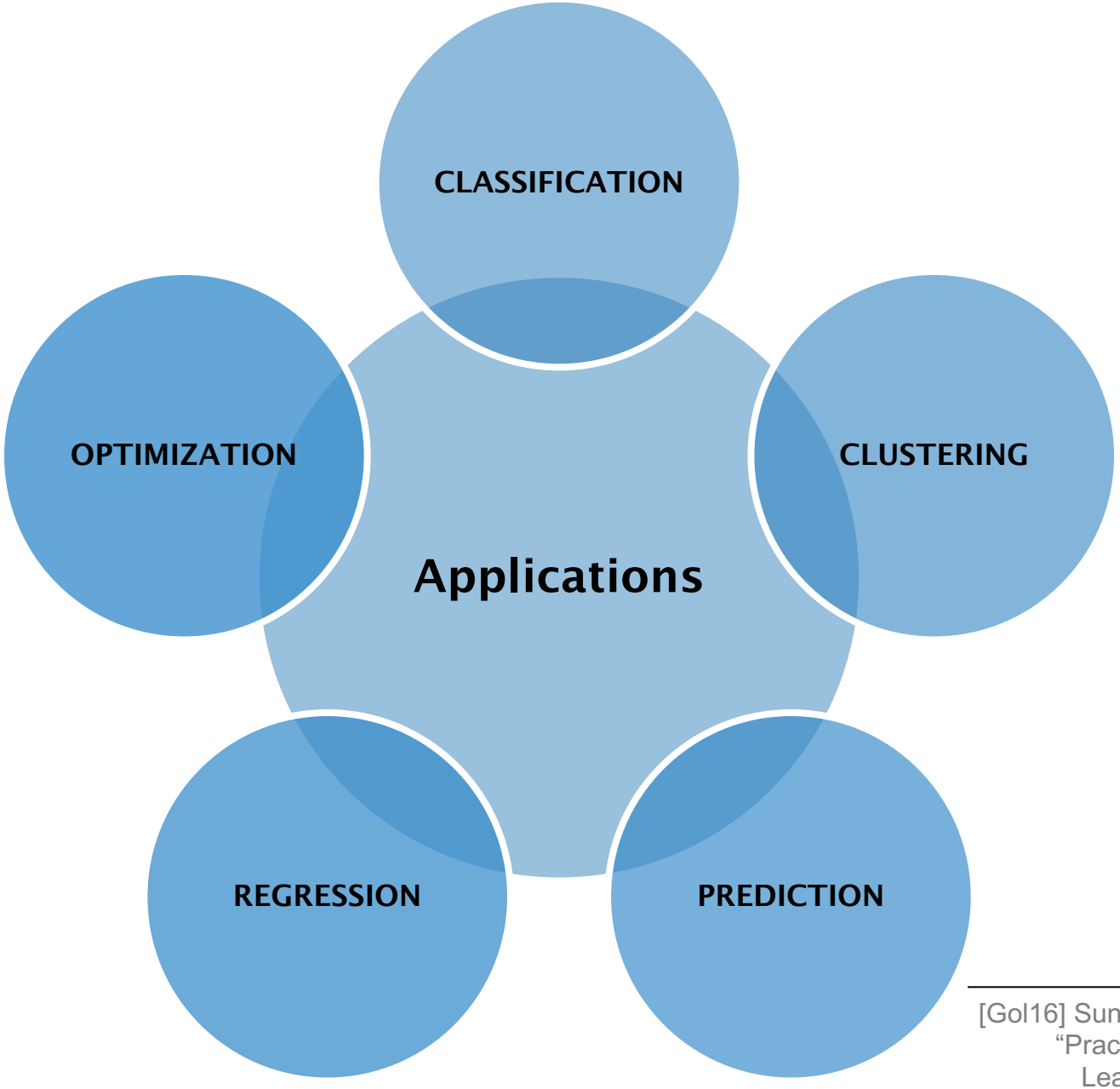
Machine Learning

Overview

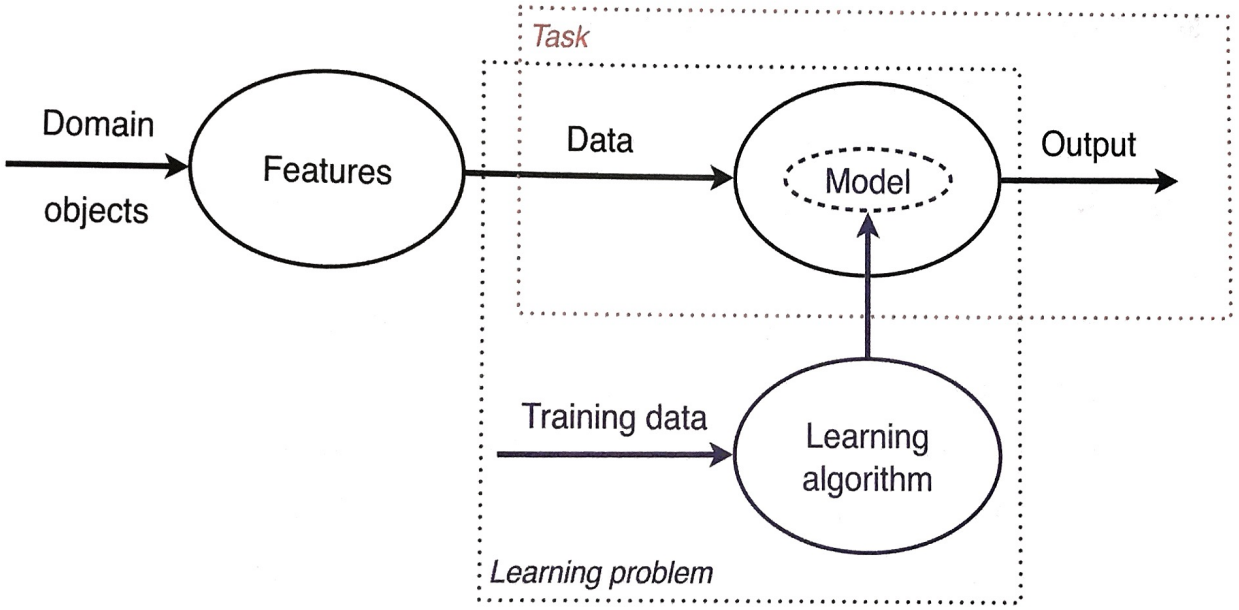
- Today we'll overview Machine Learning (ML)
- We'll consider the applications, methods as well as challenges of using ML to solve Software Engineering problems



[Gol16] Sunila Gollapudi. "Practical Machine Learning." 2016.

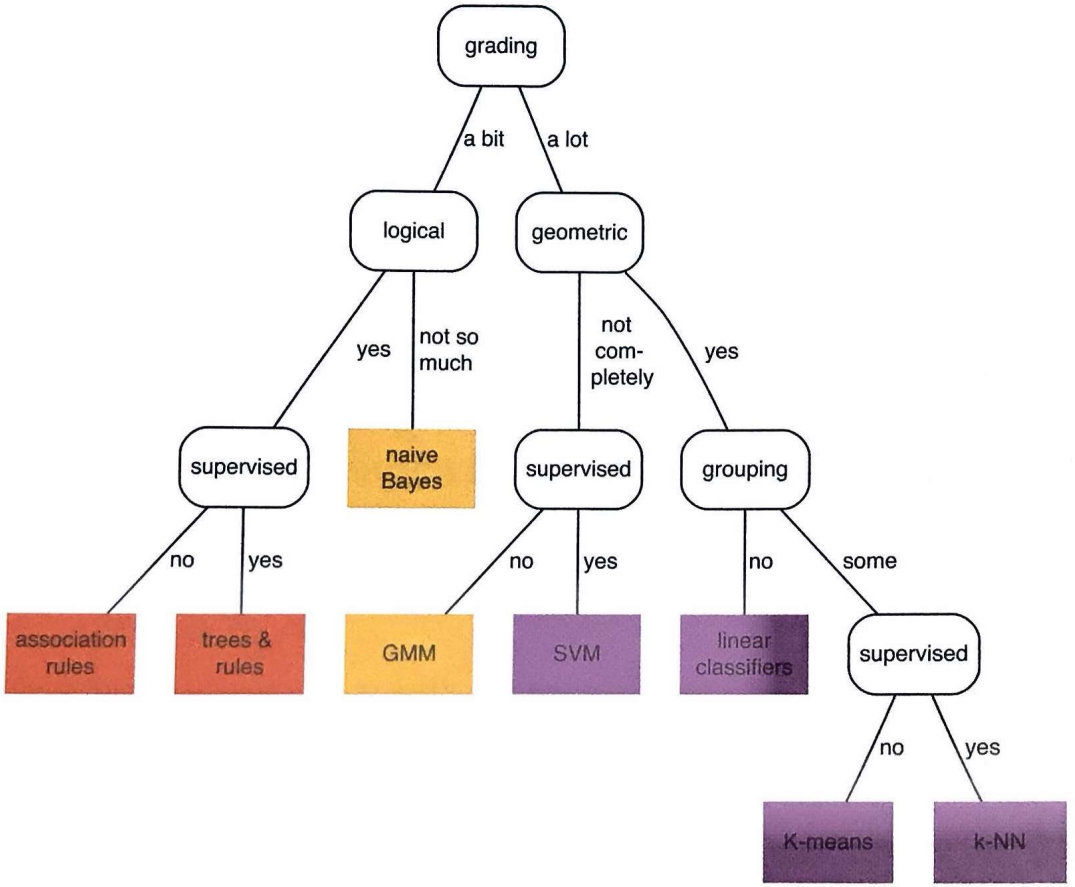
The Machine Learning (ML) Process

- Machine Learning (ML) techniques can generally be applied to tasks (problems) as follows[Fla12]:



[Fla12] Peter Flach. "Machine Learning: The Art and Science of Algorithms that Make Sense of Data." 2012.

A Machine Learning Taxonomy [Fla12]



[Fla12] Peter Flach. "Machine Learning: The Art and Science of Algorithms that Make Sense of Data." 2012.

Grading vs Grouping

- **Grouping models**
  - have a fixed/finite resolution
  - involve *"breaking up the instance space into groups or segments, the number of which is determined at training time"* [Fla12].
- **Grading models**
  - have an infinite resolution
  - *"Rather than applying very simple, local models, they form one global model over the instance space. Consequently, grading models are (usually) able to distinguish between arbitrary instances, no matter how similar they are"* [Fla12].

Features

- Features are *"functions that map from the instance space to some set of feature values called the domain of the feature"* [Fla12].
- Models are defined in terms of features!
- Example feature domains:
  - Real numbers
  - Integers
  - Booleans
  - …

Models

- **Geometric models**
  - constructed in the instance space, involves lines, planes, distance, etc.
- **Probabilistic models**
  - likelihoods, posterior probabilities
- **Logical models**
  - rule based, algorithmic, often organised in tree structure

Training in ML

- Supervised Learning
- Unsupervised Learning
- Semi-Supervised Learning

Bayesian Algorithms

Naïve Bayes

- Probabilistic classifiers
- (naïve) independence assumptions

Average One-Dependence Estimators (AODE)

- Probablistic classification learning method
- more accurate
- a little more computationally expensive
- supports incremental learning

Bayesian Belief Network (BBN)

- Probabilistic graphical model for inference and learning
- a directed acyclic graph (random variables = nodels, conditional dependences = edges)

Kernel Algorithms

Support Vector Machine (SVM)

- Supervised learning
- Applications in classification and regression
- Can be used for multi-class tasks

Linear Discriminant Analysis (LDA)

- Works when observations of independent variables are continuous values
- Can be used for multi-class tasks

Decision Tree-Based Algorithms

Chi-Square

- A statistical hypothesis test

Classification & Regression Tree (CART)

- Classification trees are used when the target variable is categorical
- Regression trees are used when the target variable is continuous – involves prediction

Random Forest

- Ensemble learning method
- Supervised learning that work by constructing decision trees
- Applications in classification and regression

Clustering Methods

K-Means

- A method of vector quantization
- Allows new data to be classified into existing clusters

Expectation Maximization (EM)

- An iterative method for finding a maximum likelihood solution

More ML Methods…

- Dimensionality reduction
- ANN – Artificial neural networks
- Ensemble methods
- Instance-Based Learning
- Regression Analysis-based Algorithms
- Association Rule-Based Learning

Challenges with ML

- **Overfitting**
  - Occurs when your model (target function) is tailored too much to past (training) data and doesn't generalise to future data points that have not yet been observed
  - You have trained your model too well!
- **Underfitting**
  - Occurs when your model is not trained well enough for modeling the past (training) data and does not generalise to future data points either
  - You haven't trained your model well enough!

Machine Learning

Summary

- Today we conducted a high-level overview of Machine Learning's applications, methods and challenges.

Up Next

- Neural networks, deep learning
- Meta-heuristic search algorithms including genetic algorithms

Machine Learning

References

[Fla12] Peter Flach. “Machine Learning: The Art and Science of Algorithms that Make Sense of Data.” 2012.

[Gol16] Sunila Gollapudi. “Practical Machine Learning.” 2016.