Artificial Intelligence and Machine Learning

Summary/Final Review

Machine Learning

- The main objective of machine learning:
  - to build models that capture the relationships in a large dataset between descriptive features and a target feature;
  - To find the association between data items in a large dataset
- Machine learning ≈ inductive learning
  - A model/rule learned by induction is not guaranteed to be correct.
  - Learning can't occur unless the learning process is biased in some way.
- There are many open questions that must be answered to successfully apply machine learning techniques.

Steps in Building a Predictive Model

- Business Understanding
  - What is the problem to be addressed? How can a prediction model address the problem? What data is available?
- Data Understanding
  - What are the domain concepts? What is the target feature? What descriptive features should be used?
- Data Preparation
  - Are there data quality issues? How do we deal with the data quality issues? What features should we include?
- Modeling
  - What types of models should we use? How to set the parameters of the machine learning algorithms? Have under fitting or overfitting occurred?
- Evaluation
  - How to measure the performance of the model? Is the model fit for the purpose?
- Deployment
  - How can the model be evaluated after the deployment? How will the model be integrated into the organization?

Different Perspectives on Prediction Models

- Information Based

$$H(t, D) = -\sum_{l \in levels(t)} (P(t = l) \times log_2(P(t = l)))$$

- Similarity Based

$$dist(\boldsymbol{q}, \boldsymbol{d}) = \sqrt{\sum_{i=1}^{m} (\boldsymbol{q}[i] - \boldsymbol{d}[i])^2}$$

- Probability Based

$$P(t = l|\boldsymbol{q}) = \frac{P(\boldsymbol{q}|t = l)P(t = l)}{P(\boldsymbol{q})}$$

- Error Based

$$L_2(M_{\boldsymbol{W}}, D) = \frac{1}{2}\sum_{i=1}^{n} (t_i - M_{\boldsymbol{W}}(\boldsymbol{d}_i))^2$$

Different Perspectives

- Parametric vs non-parametric
  - Whether the size of the domain representation used to define a model is solely determined by the number of features or is affected by the number of instances in the dataset
- Generative vs discriminative
  - A model is generative if it can be used to generate data that will have the same characteristics as the dataset from which the model was produced.
  - A discriminative model learn the boundary between classes rather than the characteristics of the distributions of the different classes.

Taxonomy of Models

Model	Parametric/ Non-Parametric	Generative/ Discriminative
K Nearest neighbor	Non-Parametric	Generative
Decision Trees	Non-Parametric	Discriminative
Naïve Bayes	Parametric	Generative
Bayesian Network	Parametric	Generative
Linear Regression	Parametric	Discriminative
Logistic Regression	Parametric	Discriminative

Choosing a Machine Learning Approach

- There is not one best approach that always outperforms the others.
- Typically choose a number of different approaches and run experiments to evaluate which one best suits the particular project.
- Two questions to consider in the selection of a set of initial approaches:
  - Does a machine learning approach match the requirements of the project?
  - Is it a classification, association or clustering task?
  - Is the approach suitable for the type of prediction and the types of descriptive features if it's a classification task?

Data Considerations

- Continuous target → error based models
- Categorical target → information and/or probability models
- Cont. descriptive features + cat. Target → similarity based models'
- Cont. descriptive features + cont. target → error based models
- Cat. Descriptive features → information and/or probability based models
- Lots of descriptive features (curse of dimensionality) → feature selection/reduction

Key Tasks in Building a Prediction Model

- Become situationally fluent so that we can converse with experts in the application domain
- Explore the data to understand it correctly
- Spend time clean the data
- Think hard about the best ways to represent features
- Spend time designing the evaluation process correctly

Learning Association Rules

- Finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transaction databases, relational databases, and other information repositories.
- Frequent patterns and Associations: Given a set of transactions, find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction

Association Rule

- **Association Rule**
  - An implication expression of the form $X \rightarrow Y$, where X and Y are itemsets
- Rule Evaluation
  - Support
  - Confidence
  - Interestingness
- The Market-Basket Model

Clustering

- What is clustering analysis
  - Grouping similar data points together
  - Unsupervised learning process
  - Minimize the dissimilarity within cluster
  - Maximize the dissimilarity with other clusters
- Applications

Major Clustering Approaches

- Partitioning
  - K-means
  - Bisecting K-means
- Hierarchical
  - Agglomerative
    - Distance between clusters
  - Divisive
    - MST based

Cluster Validation

- External
  - Entropy
  - Purity/classification error rate
- Internal
  - Sum of squared errors
  - Cohesion/separation
  - Silhouette coefficient

Outlier Detection

- What is it?
- Applications
- Working assumption
  - More normal data than outliers
- General steps
- Methods
  - Graphical & Statistical-based
  - Distance-based
    - Nearest neighbor
    - Cluster based
  - Model-based