# Describing Machine Learning Algorithms

# Describe ML Algorithms

- Describe machine learning algorithms using
  1. statistics, probability and linear algebra - mathematical descriptions are precise and unambiguous
  2. Or machine learning algorithms for developers in repeatable procedures.
- To use the developers approach: how machine learning algorithms work, not the theory behind why things work not the derivations of equations

1. In terms of the representation used by the algorithm (the actual numbers and structure that could be stored in a file).

2. In terms of the abstract repeatable procedures used by the algorithm to learn a model from training data and later to make predictions with the model.

3. The procedure used by the algorithm to make predictions given a learned model (With examples showing how real numbers plug into the equations and what numbers to expect as output)

# Data – Terminology to describe data in ML

1. Standard data terminology used in general when talking about spreadsheets of data.
2. Data terminology used in statistics and the statistical view of machine learning.
3. Data terminology used in the computer science perspective of machine learning.

# Data – Terminology to describe data in ML

1. Tabular data: This form of data is easy to work with in machine learning

- Column: A column describes data of a single type. Ex: weights, heights, prices etc. All the data in one column will have the same scale
- Row: A row describes a single entity or observation and the columns describe properties about that entity or observation. The more rows you have, the more examples from the problem domain that you have.
- Cell: A cell is a single value in a row and column. It may be a real value (1.5) an integer (2) or a category (red).

| ◇ | A | B | C | D |
|---|---|---|---|---|
| 1 | | Column 1 | Column 2 | Column 3 |
| 2 | Row 1 | 2.2 | 2.3 | 1 |
| 3 | Row 2 | 2.3 | 2.6 | 0 |
| 4 | Row 3 | 2.1 | 2 | 1 |

# Data – Terminology to describe data in ML

- Statistical Learning Perspective of data in ML

| | A | B | C |
|---|---|---|---|
| 1 | X1 | X2 | Y |
| 2 | 2.2 | 2.3 | 1 |
| 3 | 2.3 | 2.6 | 0 |
| 4 | 2.1 | 2 | 1 |
| 5 | | | |

# Data – Terminology to describe data in ML

## 2. Statistical Learning Perspective of data in ML

- The statistical perspective frames data in the context of a hypothetical function (f) that the machine learning algorithm is trying to learn.
- That is, given some input variables (input), what is the predicted output variable (output).
- Output = f(Input)
- Those columns that are the inputs are referred to as input variables.
- the column of data that we would like to predict for new input data in the future is called the output variable. It is also called the response variable.
- Output Variable = f(Input Variables)

# Data – Terminology to describe data in ML

## 2. Statistical Learning Perspective of data in ML

- If we have more than one input variable, the group of input variables are referred to as the input vector.
- Output Variable = f(Input Vector)
- input variables referred as  independent variables
- output variable as the dependent variable.
- Phrasing of the prediction problem the output is dependent or a function of the input or independent variables.
- Dependent Variable = f(Independent Variables)

# Data – Terminology to describe data in ML

## 2. Statistical Learning Perspective of data in ML

- The standard shorthand used in the statistical perspective is to refer to the input variables as X and the output variables as Y.
- $Y = f(X)$
- When we have multiple input variables they may be dereferenced with an integer to indicate their ordering in the input vector, for example $X1$, $X2$ and $X3$ for data in the first three columns.

# Data – Terminology to describe data in ML

## 3. Computer Science Perspective of Data in ML

- A row describes an entity (like a person) or an observation about an entity.
- the columns for a row are often referred to as attributes of the observation.
- When modeling a problem and making predictions, we may refer to input attributes and output attributes.
- OutputAttribute = Program(InputAttributes)

# Data – Terminology to describe data in ML

## 3. Computer Science Perspective of Data in ML

- Another name for columns is features: common when working with data where features must be extracted from the raw data in order to construct an observation.
- Examples: analog data like images, audio and video.
- Output = Program(InputFeatures)
- Another computer science phrasing: row of data as an instance
- a row may be considered a single example or single instance of data observed or generated by the problem domain.
- Prediction = Program(Instance)

# Data – Terminology to describe data in ML

- Models and Algorithms
- Think of the model as the specific representation learned from data and the algorithm as the process for learning it.
- Model = Algorithm(Data)
- Ex: a decision tree or a set of coefficients are a model
- C5.0 and Least Squares Linear Regression are algorithms to learn those respective models.

# Algorithms Learn a Mapping From Input to Output

- How do machine learning algorithms work?
- There is a common principle that underlies all supervised machine learning algorithms for predictive modeling.

# Learning a Function

- Machine learning algorithms are described as learning a target function (f) that best maps input variables (X) to an output variable (Y ).
- Y = f(X)
- This is a general learning task where we would like to make predictions in the future (Y ) given new examples of input variables (X).
- We don't know what the function (f) looks like or it's form.
- If we did, we would use it directly and we would not need to learn it from data using machine learning algorithms.
- There is also error (e) that is independent of the input data (X).
- Y = f(X) + e

# Learning a Function

- This error might be
  - Error such as not having enough attributes to sufficiently characterize the best mapping from X to Y .
- This error is called irreducible error because no matter how good we get at estimating the target function (f), we cannot reduce this error.

- Note: This is to say, that the problem of learning a function from data is a difficult problem
- This is the reason why the field of machine learning and machine learning algorithms exist.

# Techniques For Learning a Function- Summary

- The most common type of machine learning is to learn the mapping $Y = f(X)$ to make predictions of Y for new X. This is called predictive modeling or predictive analytics
- Goal is to make the most accurate predictions possible.
- When we learn a function (f) we are estimating its form from the data that we have available.
  - this estimate will have error. It will not be a perfect estimate for the underlying hypothetical best mapping from Y given X.
- Much time in applied machine learning is spent attempting
  - to improve the estimate of the underlying function and
  - To improve the performance of the predictions made by the model

# Techniques For Learning a Function- Summary

- Machine learning algorithms are techniques for estimating the target function (f) to predict the output variable (Y ) given input variables (X).
- Different representations
  - make different assumptions about the form of the function being learned
    - such as whether it is linear or nonlinear.
  - make different assumptions about the shape and structure of the function and
  - how best to optimize a representation to approximate it.
- So it is so important to try a suite of different algorithms on a machine learning problem,
- because we cannot know before hand which approach will be best at estimating the structure of the underlying function we are trying to approximate.

# Parametric and Nonparametric Machine Learning Algorithms

- Assumptions can greatly simplify the learning process, but can also limit what can be learned.
- Algorithms that simplify the function to a known form are called parametric machine learning algorithms.
- A learning model that summarizes data with a set of parameters of fixed size (independent of the number of training examples) is called a parametric model.
- No matter how much data thrown at a parametric model, no change about how many parameters it needs

# Parametric Machine Learning Algorithms

Parametric machine learning algorithms involve two steps:

1. Select a form for the function.
2. Learn the coefficients for the function from the training data.

- An easy to understand functional form for the mapping function is a line, as is used in linear regression:
- B0 + B1 * X1 + B2 * X2 = 0
- Where B0, B1 and B2 are the coefficients of the line that control the intercept and slope, and X1 and X2 are two input variables.
- Assuming the functional form of a line greatly simplifies the learning process.
- Now, all we need to do is estimate the coefficients of the line equation and we have a predictive model for the problem.

# Parametric Machine Learning Algorithms

- Often the assumed functional form is a linear combination of the input variables
- parametric machine learning algorithms are often also called linear machine learning algorithms.
- Problem: the actual unknown underlying function may not be a linear function like a line.
  - It could be almost a line and require some minor transformation of the input data to work right.
  - Or it could be nothing like a line in which case the assumption is wrong and the approach will produce poor results.

# Parametric Machine Learning Algorithms

## Examples of parametric machine learning algorithms

- Logistic Regression
- Linear Discriminant Analysis
- Perceptron

## Benefits of Parametric Machine Learning Algorithms:

- Simpler: easier to understand and interpret results.
- Speed: Parametric models are very fast to learn from data.
- Less data: They do not require as much training data and can work well even if the fit to the data is not perfect.

## Limitations of Parametric Machine Learning Algorithms:

- Constrained: By choosing a functional form these methods are highly constrained to the specified form.
- Limited Complexity: The methods are more suited to simpler problems.
- Poor fit: In practice the methods are unlikely to match the underlying mapping function.

# Nonparametric Machine Learning Algorithms

- Algorithms that do not make strong assumptions about the form of the mapping function are called nonparametric machine learning algorithms.
- By not making assumptions, they are free to learn any functional form from the training data.
- Nonparametric methods are good when we have a lot of data and no prior knowledge
- Nonparametric methods seek to best fit the training data in constructing the mapping function, whilst maintaining some ability to generalize to unseen data.
- They are able to fit a large number of functional forms.
- Ex: k-nearest neighbors algorithm makes predictions based on the k most similar training patterns for a new data instance.
  - The method does not assume anything about the form of the mapping function

# Nonparametric Machine Learning Algorithms

- Examples of popular nonparametric machine learning algorithms
- Decision Trees like CART and C4.5
- Naive Bayes
- Support Vector Machines
- Neural Networks
- Benefits of Nonparametric Machine Learning Algorithms:
- Flexibility: Capable of fitting a large number of functional forms.
- Powerful: No assumptions (or weak assumptions) about the underlying function.
- Performance: Can result in higher performance models for prediction.
- Limitations of Nonparametric Machine Learning Algorithms:
- More data: Require a lot more training data to estimate the mapping function.
- Slower: A lot slower to train as they often have far more parameters to train.
- Overfitting: More of a risk to overfit the training data and it is harder to explain why specific predictions are made.

# Supervised Machine Learning Algorithms

- Supervised learning:
- We have input variables (X) and an output variable (Y ) and use an algorithm to learn the mapping function from the input to the output.
- Y = f(X)
- The goal is to approximate the mapping function that when we have new input data (X) we can predict the output variables (Y ) for that data.
- The process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process.
- We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher.
- Learning stops when the algorithm achieves an acceptable level of performance.

# Supervised Machine Learning Algorithms

- Grouped into regression and classification problems.
- Classification: A classification problem is when the output variable is a category, such as red or blue or disease and no disease.
- Regression: A regression problem is when the output variable is a real value, such as height, or weight.

# Supervised Machine Learning Algorithms

- Types of problems built on top of classification and regression include
  - recommendation and time series prediction respectively.
- Examples of supervised machine learning algorithms are:
  - Linear regression for regression problems.
  - Random forest for classification and regression problems.
  - Support vector machines for classification problems.

# Unsupervised Machine Learning Algorithms

- Unsupervised learning is we only have input data (X) and no corresponding output variables.
- The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.
- These are called unsupervised learning because unlike supervised learning above there is no correct answers and there is no teacher.
- Algorithms are left to discover and present the interesting structure in the data.
- Unsupervised learning problems can be further grouped into clustering and association problems.

# Unsupervised Machine Learning Algorithms

- Clustering: A clustering problem is when we want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
- Association: An association rule learning problem is when we want to discover rules that describe large portions of data, such as people that buy A also tend to buy B.
- Examples of unsupervised learning algorithms are:
  - K-means for clustering problems.
  - Apriori algorithm for association rule learning problems.