AI: Strategy + Marketing (MGT 853)

ML Essentials (Session 2)

Vineet Kumar

Yale School of Management

Spring 2024

Definitions: AI and ML

Artificial Intelligence

"...Intelligence can in principle be so precisely described that a machine can be made to simulate it." (John McCarthy)

Machine Learning

- "The field of study that gives computers the ability to learn without explicitly being programmed" (Arthur Samuel)

What's the difference?

Definitions: AI and ML

Artificial Intelligence

"...Intelligence can in principle be so precisely described that a machine can be made to simulate it." (John McCarthy)

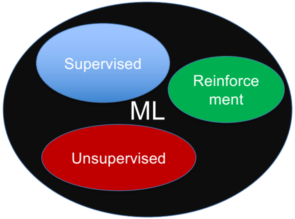
Machine Learning

- "The field of study that gives computers the ability to learn without explicitly being programmed" (Arthur Samuel)
- "Improve over Task T with resect to some performance measure P based on experience E" (Tom Mitchell)

What's the difference?

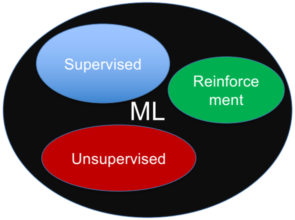
Differences between ML and other AI approaches

- ML primarily is "learning from data"



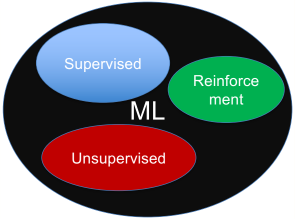
Differences between ML and other AI approaches

- ML primarily is "learning from data"
- Human domain knowledge is *not* required or even expected



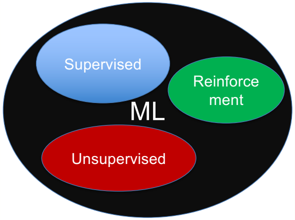
Differences between ML and other AI approaches

- ML primarily is "learning from data"
- Human domain knowledge is *not* required or even expected
- Same algorithm can be used for very different applications (e.g. cancer detection or astronomy or cats /dogs)



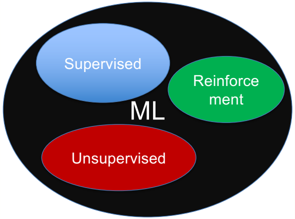
Differences between ML and other AI approaches

- ML primarily is "learning from data"
- Human domain knowledge is *not* required or even expected
- Same algorithm can be used for very different applications (e.g. cancer detection or astronomy or cats /dogs)
- Since 2010+ ML, especially Deep Learning, has dominated other approaches when measured by accuracy



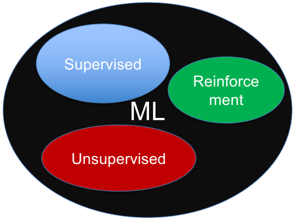
Differences between ML and other AI approaches

- ML primarily is "learning from data"
- Human domain knowledge is *not* required or even expected
- Same algorithm can be used for very different applications (e.g. cancer detection or astronomy or cats /dogs)
- Since 2010+ ML, especially Deep Learning, has dominated other approaches when measured by accuracy



Differences between ML and other AI approaches

- ML primarily is "learning from data"
- Human domain knowledge is *not* required or even expected
- Same algorithm can be used for very different applications (e.g. cancer detection or astronomy or cats /dogs)
- Since 2010+ ML, especially Deep Learning, has dominated other approaches when measured by accuracy



Rest of the course will focus on ML (use interchangeably with AI)

Summary of Session 1

- AI not a new technology, has been around

Summary of Session 1

- AI not a new technology, has been around
- Recent trends in algorithms, data and computing have converged to make AI more valuable

Summary of Session 1

- AI not a new technology, has been around
- Recent trends in algorithms, data and computing have converged to make AI more valuable
- Currently ML $\approx$ AI, since classical AI not as common

Summary of Session 1

- AI not a new technology, has been around
- Recent trends in algorithms, data and computing have converged to make AI more valuable
- Currently ML $\approx$ AI, since classical AI not as common
- ML is focused on prediction and "learning from data" using many different methods

Summary of Session 1

- AI not a new technology, has been around
- Recent trends in algorithms, data and computing have converged to make AI more valuable
- Currently ML $\approx$ AI, since classical AI not as common
- ML is focused on prediction and "learning from data" using many different methods
- ML: Supervised, Unsupervised and Reinforcement

Summary of Session 1

- AI not a new technology, has been around
- Recent trends in algorithms, data and computing have converged to make AI more valuable
- Currently ML $\approx$ AI, since classical AI not as common
- ML is focused on prediction and "learning from data" using many different methods
- ML: Supervised, Unsupervised and Reinforcement
    - Designed to answer different types of questions, data different

Summary of Session 1

- AI not a new technology, has been around
- Recent trends in algorithms, data and computing have converged to make AI more valuable
- Currently ML $\approx$ AI, since classical AI not as common
- ML is focused on prediction and "learning from data" using many different methods
- ML: Supervised, Unsupervised and Reinforcement
    - Designed to answer different types of questions, data different
- Integrating Prediction with decision making can be challenging (Module B)

Supervised Learning

- Trying to predict some variable y based on other data X

Supervised Learning

- Trying to predict some variable y based on other data X
- y: label / target / output / dependent variable

Supervised Learning

- Trying to predict some variable y based on other data X
- y: label / target / output / dependent variable
- X: predictor, covariate, explanatory variable

Supervised Learning

- Trying to predict some variable y based on other data X
- y: label / target / output / dependent variable
- X: predictor, covariate, explanatory variable
- Represent this as $\boxed{y = f(X)}$

Supervised Learning

- Trying to predict some variable y based on other data X
- y: label / target / output / dependent variable
- X: predictor, covariate, explanatory variable
- Represent this as $\boxed{y = f(X)}$
- Do Humans learn like this?

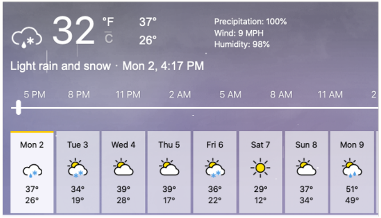
Supervised Learning

- Trying to predict some variable y based on other data X
- y: label / target / output / dependent variable
- X: predictor, covariate, explanatory variable
- Represent this as $\boxed{y = f(X)}$
- Do Humans learn like this?

Supervised Learning

- Trying to predict some variable y based on other data X
- y: label / target / output / dependent variable
- X: predictor, covariate, explanatory variable
- Represent this as $\boxed{y = f(X)}$
- Do Humans learn like this?

Most commonly used form of ML in practice



Logistic Regression
Polynomial Regression
Support Vector Machines
Decision Trees
Deep Neural Nets

Unsupervised Learning

- Trying to identify patterns in the data

Unsupervised Learning

- Trying to identify patterns in the data
- We do not have a output label y only X variables

Unsupervised Learning

- Trying to identify patterns in the data
- We do not have a output label y only X variables
- Cannot represent this as with structured learning

Unsupervised Learning

- Trying to identify patterns in the data
- We do not have a output label y only X variables
- Cannot represent this as with structured learning
- Not trying to make any prediction here

Unsupervised Learning

- Trying to identify patterns in the data
- We do not have a output label y only X variables
- Cannot represent this as with structured learning
- Not trying to make any prediction here

Unsupervised Learning

- Trying to identify patterns in the data
- We do not have a output label y only X variables
- Cannot represent this as with structured learning
- Not trying to make any prediction here

Can use for exploratory analysis and segmentation even when question is unclear

Cluster Analysis
K-means
K-Nearest Neighbor
Association Rule Mining
Principal Components Analysis

Agenda for Today's Session

- Impact of AI

Agenda for Today's Session

- Impact of AI
- Elements of ML model

Agenda for Today's Session

- Impact of AI
- Elements of ML model
    - Supervised Models

Agenda for Today's Session

- Impact of AI
- Elements of ML model
  - Supervised Models
  - Unsupervised Models

Agenda for Today's Session

- Impact of AI
- Elements of ML model
    - Supervised Models
    - Unsupervised Models

Agenda for Today's Session

- Impact of AI
- Elements of ML model
    - Supervised Models
    - Unsupervised Models

Agenda for Today's Session

- Impact of AI
- Elements of ML model
  - Supervised Models
  - Unsupervised Models

Practicum

- Intro to Google Colab
- Building an ML model in Colab
- Before the model (Pre-processing)

Agenda for Today's Session

- Impact of AI
- Elements of ML model
    - Supervised Models
    - Unsupervised Models

Practicum

- Intro to Google Colab
- Building an ML model in Colab
- Before the model (Pre-processing)
- In the model (In-processing)

Agenda for Today's Session

- Impact of AI
- Elements of ML model
  - Supervised Models
  - Unsupervised Models

Practicum

- Intro to Google Colab
- Building an ML model in Colab
- Before the model (Pre-processing)
- In the model (In-processing)
- After the Model (Post-processing)

Building Blocks of ML models

Data

Training, Validation and Test

Why do we need to split data?

Peeking is a consequence of using test-set performance to both choose a hypothesis and evaluate it. The way to avoid this is to really hold the test set out – lock it away until you are completely done with learning and simply wish to obtain an independent evaluation of the final hypothesis. (And then, if you don't like the results ... you have to obtain, and lock away, a completely new test set if you want to go back and find a better hypothesis.) – **Russell and Norvig (Artificial Intelligence: A Modern Approach)**

Data

Training, Validation and Test

Why do we need to split data?

If the test set is locked away, but you still want to measure performance on unseen data as a way of selecting a good hypothesis, then divide the available data (without the test set) into a training set and a validation set. – **Russell and Norvig (Artificial Intelligence: A Modern Approach)**

Pre-processing

Why?

What is Data Pre-processing?

Transforming the Raw Data to a format that is suitable for use in an ML algorithm. Two types of pre-processing: Lossy and Lossless

- Identify anomalies in the data (missing values, duplicates, outliers, inconsistent or implausible values etc.)

Pre-processing

Why?

What is Data Pre-processing?

Transforming the Raw Data to a format that is suitable for use in an ML algorithm. Two types of pre-processing: Lossy and Lossless

- Identify anomalies in the data (missing values, duplicates, outliers, inconsistent or implausible values etc.)
- Convenient format for input

Pre-processing

Why?

What is Data Pre-processing?

Transforming the Raw Data to a format that is suitable for use in an ML algorithm. Two types of pre-processing: Lossy and Lossless

- Identify anomalies in the data (missing values, duplicates, outliers, inconsistent or implausible values etc.)
- Convenient format for input
- Dimensionality Reduction

Pre-processing

Why?

What is Data Pre-processing?

Transforming the Raw Data to a format that is suitable for use in an ML algorithm. Two types of pre-processing: Lossy and Lossless

- Identify anomalies in the data (missing values, duplicates, outliers, inconsistent or implausible values etc.)
- Convenient format for input
- Dimensionality Reduction
- Consistent format of possibly acquired from multiple sources

Pre-processing

Why?

What is Data Pre-processing?

Transforming the Raw Data to a format that is suitable for use in an ML algorithm. Two types of pre-processing: Lossy and Lossless

- Identify anomalies in the data (missing values, duplicates, outliers, inconsistent or implausible values etc.)
- Convenient format for input
- Dimensionality Reduction
- Consistent format of possibly acquired from multiple sources
- Reduce algorithmic bias (How?)

Feature Engineering

What is Feature Engineering?

The process of identifying specific features, transforming them, or *creating new features* from data is called feature engineering.

- “some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used.” – Pedro Domingos

Feature Engineering

What is Feature Engineering?

The process of identifying specific features, transforming them, or *creating new features* from data is called feature engineering.

- “some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used.” – Pedro Domingos
- Land price (Price) based on rectangular plot characteristics: length (L) and width (W)

Feature Engineering

What is Feature Engineering?

The process of identifying specific features, transforming them, or *creating new features* from data is called feature engineering.

- "some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used." – Pedro Domingos
- Land price (Price) based on rectangular plot characteristics: length (L) and width (W)
  - $Price_i = \beta_0 + \beta_1 L_i + \beta_2 W_i$

Feature Engineering

What is Feature Engineering?

The process of identifying specific features, transforming them, or *creating new features* from data is called feature engineering.

- "some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used." – Pedro Domingos
- Land price (Price) based on rectangular plot characteristics: length (L) and width (W)
  - $Price_i = \beta_0 + \beta_1 L_i + \beta_2 W_i$
  - Is this linear model sufficient? What about a quadratic model?

Feature Engineering

What is Feature Engineering?

The process of identifying specific features, transforming them, or *creating new features* from data is called feature engineering.

- "some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used." – Pedro Domingos
- Land price (Price) based on rectangular plot characteristics: length (L) and width (W)
    - $Price_i = \beta_0 + \beta_1 L_i + \beta_2 W_i$
    - Is this linear model sufficient? What about a quadratic model?
- Lots of possibilities with unstructured data

Hyperparameters

What are Hyperparameters?

These are parameters typically not learned from the data during training, but rather set to specific values when implementing.

- Also called tuning parameters.

Hyperparameters

What are Hyperparameters?

These are parameters typically not learned from the data during training, but rather set to specific values when implementing.

- Also called tuning parameters.
- Control the operation of the learning algorithm

Hyperparameters

What are Hyperparameters?

These are parameters typically not learned from the data during training, but rather set to specific values when implementing.

- Also called tuning parameters.
- Control the operation of the learning algorithm
- K-means: ??

Hyperparameters

What are Hyperparameters?

These are parameters typically not learned from the data during training, but rather set to specific values when implementing.

- Also called tuning parameters.
- Control the operation of the learning algorithm
- K-means: ??
- Set based on application and purpose of ML algorithm

Hyperparameters

What are Hyperparameters?

These are parameters typically not learned from the data during training, but rather set to specific values when implementing.

- Also called tuning parameters.
- Control the operation of the learning algorithm
- K-means: ??
- Set based on application and purpose of ML algorithm
- For more: https://towardsdatascience.com/the-what-why-and-how-of-hyperparameter-tuning-for-machine-learning

Hyperparameters

What are Hyperparameters?

These are parameters typically not learned from the data during training, but rather set to specific values when implementing.

- Also called tuning parameters.
- Control the operation of the learning algorithm
- K-means: ??
- Set based on application and purpose of ML algorithm
- For more: `https://towardsdatascience.com/the-what-why-and-how-of-hyperparameter-tuning-for-machine-learning`
- For Random Forest hyperparameters, see: `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html`

Model Complexity

Complex models explain data better

- We create categorize models in terms of their "complexity"

Model Complexity

Complex models explain data better

- We create categorize models in terms of their "complexity"
- More complex models could have:

Model Complexity

Complex models explain data better

- We create categorize models in terms of their "complexity"
- More complex models could have:
  - more parameters

Model Complexity

Complex models explain data better

- We create categorize models in terms of their "complexity"
- More complex models could have:
  - more parameters
  - more features

Model Complexity

Complex models explain data better

- We create categorize models in terms of their "complexity"
- More complex models could have:
  - more parameters
  - more features
  - more flexible functions (quadratic versus linear)

Model Complexity

Complex models explain data better

- We create categorize models in terms of their "complexity"
- More complex models could have:
  - more parameters
  - more features
  - more flexible functions (quadratic versus linear)
- Complex models can often explain training data very well (technical video)

Model Complexity

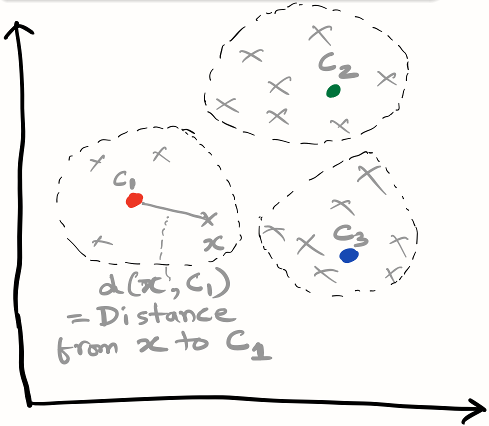
Complex models explain data better

- We create categorize models in terms of their "complexity"
- More complex models could have:
  - more parameters
  - more features
  - more flexible functions (quadratic versus linear)
- Complex models can often explain training data very well (technical video)
- But the question is do they "generalize" such performance when

K-means

What is K-means?

Group similar data together



- Group data so that points are close to the center of their clusters, but far away from other clusters
- Let's say the center of cluster i is C_i
- Any point x in the cluster has a distance of $d(x, C_i) = (x - C_i)$ from the center
- How do you decide which cluster a point "belongs to"? See which center is the closest to the point.
- There are many such points x in the cluster so we sum over all of them

K-means

- Probably the most common unstructured ML method used
- Typically used for customer segmentation (finding "similar" groups of customers)
- K is a **hyperparameter** chosen based on the business application
    - Typically $K = 2, 3, 4$ in applications
- Can devise different interventions for each group
- Other Applications:
  `https://dzone.com/articles/10-interesting-use-cases-for-the-k-means-algorithm`

Decision Trees and Random Forest

Desirable Properties of Models

Summary of Session 2

- ML is focused on prediction and “learning from data”

Summary of Session 2

- ML is focused on prediction and "learning from data"
- We have examined the elements of an ML model – consider each when building your ML pipeline

Summary of Session 2

- ML is focused on prediction and "learning from data"
- We have examined the elements of an ML model – consider each when building your ML pipeline
- Results very much depend on the quality and quantity of data

Summary of Session 2

- ML is focused on prediction and "learning from data"
- We have examined the elements of an ML model – consider each when building your ML pipeline
- Results very much depend on the quality and quantity of data
  - Think about trying to predict temperature for year using data from Jan 1 - June 30

Summary of Session 2

- ML is focused on prediction and "learning from data"
- We have examined the elements of an ML model – consider each when building your ML pipeline
- Results very much depend on the quality and quantity of data
  - Think about trying to predict temperature for year using data from Jan 1 - June 30
- ML methods have many challenges including underfitting / overfitting

Summary of Session 2

- ML is focused on prediction and "learning from data"
- We have examined the elements of an ML model – consider each when building your ML pipeline
- Results very much depend on the quality and quantity of data
  - Think about trying to predict temperature for year using data from Jan 1 - June 30
- ML methods have many challenges including underfitting / overfitting
- Need to carefully choose appropriate methods for your problem

Challenges with ML

- Model – Can I convert a problem to prediction?

Challenges with ML

- Model – Can I convert a problem to prediction?
- Data: Spurious Correlation – Is it real?

Challenges with ML

- Model – Can I convert a problem to prediction?
- Data: Spurious Correlation – Is it real?
- Generalizability

Challenges with ML

- Model – Can I convert a problem to prediction?
- Data: Spurious Correlation – Is it real?
- Generalizability
- Causality – Counterfactual Validity

Challenges with ML

- Model – Can I convert a problem to prediction?
- Data: Spurious Correlation – Is it real?
- Generalizability
- Causality – Counterfactual Validity
- Statistical Properties

Challenges with ML

- Model – Can I convert a problem to prediction?
- Data: Spurious Correlation – Is it real?
- Generalizability
- Causality – Counterfactual Validity
- Statistical Properties
- Explainability – Black Box

Challenges with ML

- Model – Can I convert a problem to prediction?
- Data: Spurious Correlation – Is it real?
- Generalizability
- Causality – Counterfactual Validity
- Statistical Properties
- Explainability – Black Box
- Fairness and Bias

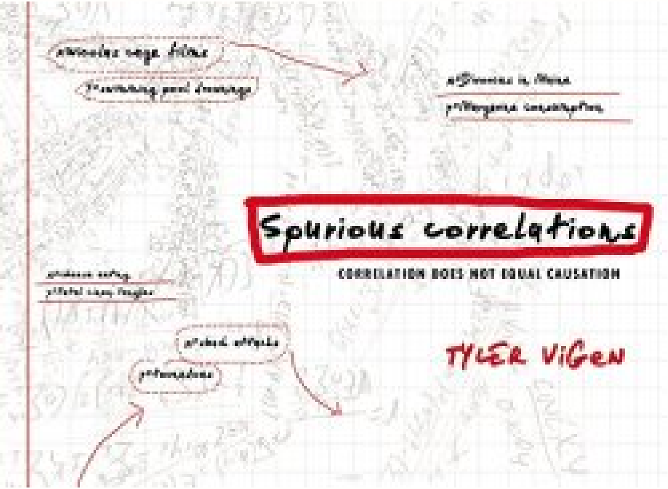
Challenges with ML

- Model – Can I convert a problem to prediction?
- Data: Spurious Correlation – Is it real?
- Generalizability
- Causality – Counterfactual Validity
- Statistical Properties
- Explainability – Black Box
- Fairness and Bias
- Moving Target – What new (better) methods are being developed?

Challenges with ML

- Model – Can I convert a problem to prediction?
- Data: Spurious Correlation – Is it real?
- Generalizability
- Causality – Counterfactual Validity
- Statistical Properties
- Explainability – Black Box
- Fairness and Bias
- Moving Target – What new (better) methods are being developed?
- Fragility - When and how does it break?

Spurious Correlations



Next Class: Deep Learning and Reinforcement Learning

- Readings for Next Class:
- Try to read Chapter 6 of Algorithms book and RL book Chapter 1 (parts 1.1 to 1.4)
- Can skim / skip Chapter 1 and 2 of Deep Learning book
- Skip over parts that are difficult – just need a general idea