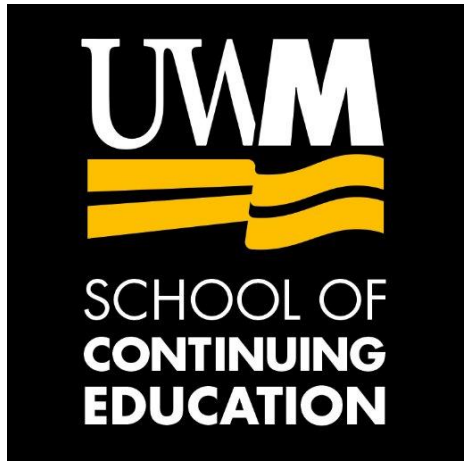


Predictive Analytics

Introduction

- Welcome from UWM
- About Corey
- Course Schedule

Live In Person Courses

- We will take breaks roughly every hour to hour and a half.
- Does anybody know others in the class
- I will be using examples from Health Care and Education.
- **DO NOT USE ANYTHING THAT SHOULD BE PRIVATE**
- I will be making my presentation available at the end of the course. **Don't take feverish notes**

Questions before we start?

Discussion Board

1. Fill out Introduction questions on discussion board
2. Reply to at least 2 other people

Data Science Tools

UWM
SCHOOL OF
CONTINUING
EDUCATION

Anaconda (Conda)

- Easiest way to perform Python/R data science and machine learning on a single machine.
- Open source
- Easy to install over 7,500 data science and machine learning packages.
- Conda makes it easy to manage multiple data environments that can be maintained and run separately without interference from each other.
- Anaconda Navigator is a desktop GUI that makes it easy to launch applications and manage packages and environments without using command-line commands.

R and Python

- Free
- Open Source
- Libraries
- Coding
- Huge network of people
- Can make data analysis, predictive analytics, reporting and visualization easier and more efficient
- Can be used for other coding/automation, but our focus will be on using them with data

Jupyter Notebooks and R Markdowns

- Open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.

- Jupyter has support for over 40 programming languages, including Julia, Python, R and Scala. R Markdown mainly support R.

- Can produce rich, interactive output: HTML, images, videos

JupyterLab and R Studio

- JupyterLab makes Python easier to use and RStudio makes R easier to use.
- Open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.
- You can arrange multiple documents and activities side by side in the work area using tabs and splitters.
- It includes a console, syntax-highlighting editor that supports direct code editing and execution, as well as tools for plotting, history, debugging and workspace management.

GitHub

- Software development platform
- Host and review code
  - Version control of code
  - Sharing of code
- Manage projects
- Huge community of open-source projects

Assignment

- If using remote desktop
  - Login
  - Search for Anaconda Navigator to make sure it's available
- If using your own computer
  - Install Anaconda
    - Install for: Just Me
    - Leave all default settings
    - Uncheck the boxes after installing about viewing documentation

Demo Together

- Anaconda Navigator
- Anaconda Command Prompt to install Git
  - conda install git
  - Type Y when asked
- JupyterLab
  - Terminal to Clone Git Repository
    - git clone https://github.com/fritschcm272/UWM_SCE_Predictive_Analytics.git
  - Command Prompt For Python Packages
    - pip install pycaret-ts-alpha==3.0.0.dev1649017462
    - pip install ydata-profiling –ignore-installed –user
    - pip install scikit-learn==1.0.2

Break Time

- Meet back at 10:55 AM
- Let me know if you have not gotten Git Repo cloned into whichever version of Anaconda you are using

Introduction to Predictive Analytics

UWM
SCHOOL OF
CONTINUING
EDUCATION

Types of Analytics

- Descriptive Analytics: Summarizes historical to better understand what is going on or what has happened.
- Diagnostic Analytics: Focus on past performance to determine what happened and why.
- Predictive Analytics: Predicting the possible outcome of future data.
- Prescriptive Analytics: Type of predictive analytics that is used to drive action.

Predictive Model Framework

- Do you need a predictive model?
- Do you have a predictive model built?
- Does your predictive model perform as expected on relevant population?
- Do you have an implementation and workflow strategy?

Predictive Analytics

- Take your knowledge of data and turn it into meaningful insights
- Help you not only understand what your previously collected data shows, but what is most likely to happen in the future
- Combination of Art and Science

Predictive Model - an algorithm created using retrospective data for predicting the likelihood of an event

Predictive Analytics Variables

- Dependent Variable (Target)
  - What you want to predict
- Independent Variables (Features)
  - What you use to predict

Predictive Analytics

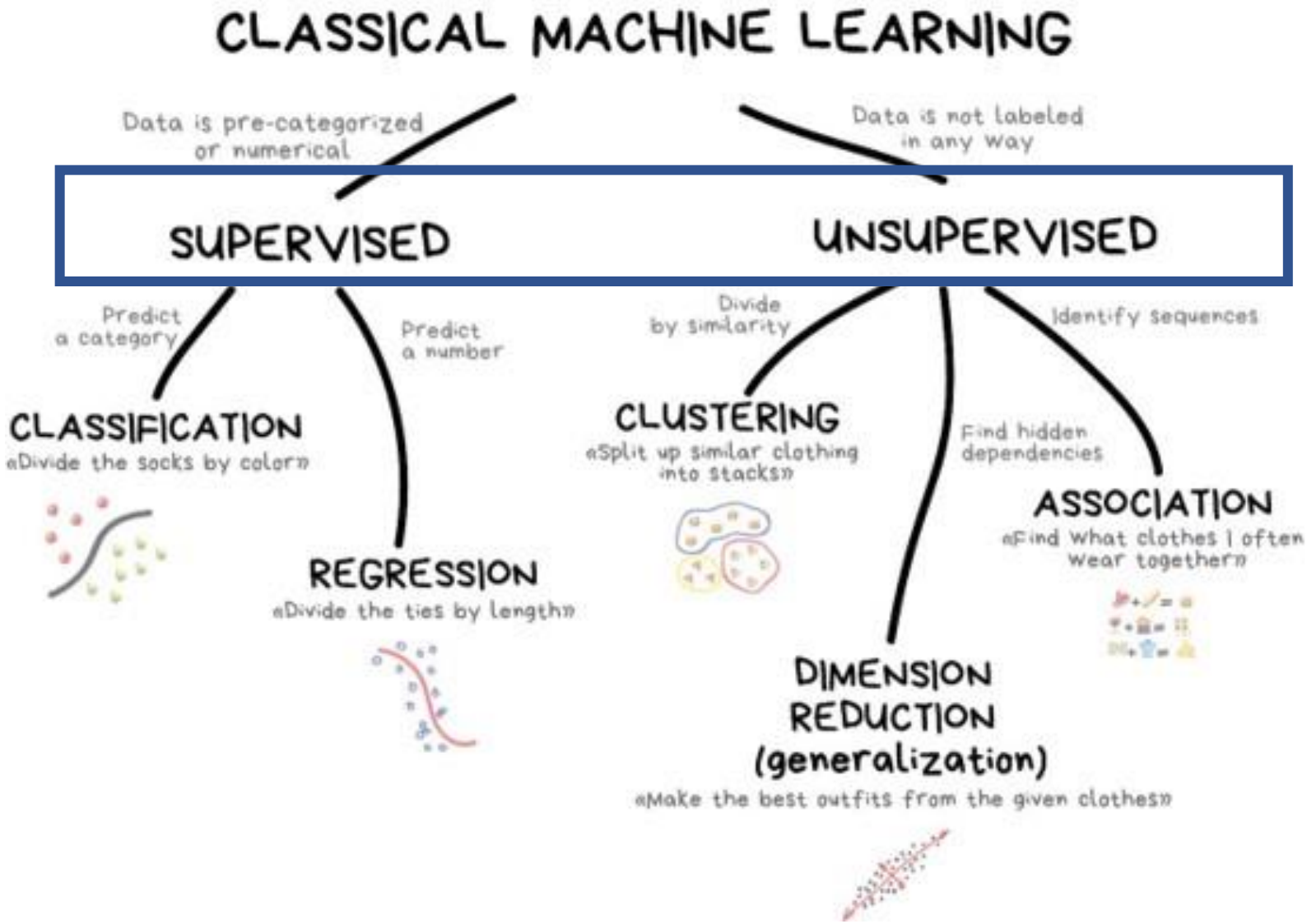
- Use previous data to find trends and patterns, to predict future outcomes
- Tries to explain and outperform randomness
- Software packages like R and Python are great with predictive models. Excel, not so much.
- Done through building statistical or machine learning predictive models

Statistical Model - mathematical model that embodies a set of statistical assumptions concerning the generation of sample data (and similar data from a larger population)

Machine Learning - Branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

Machine Learning

- Machine learning gives computers the ability to learn without being explicitly programmed. (Arthur Samuel, 1959)
- Teaching computers how to learn patterns from data to make decisions or predictions
- Blend of discovery, engineering, and business
- One of the major issues with a vast amount of data is that it can become extremely noisy and hard to understand what is truly meaningful related to these problems.
- Supervised, Unsupervised, Semi-Supervised, and Reinforcement Learning



Supervised Machine Learning

- Learning from labeled data using classification and regression models.

- Your email provider kindly places that sketchy email from the “Nigerian prince with $50,000 to deposit into an overseas bank account” into the spam folder.

Unsupervised Machine Learning

- Learning from unlabeled data using factor and cluster analysis models.

- Marketing firms “kindly” use hundreds of behavior and demographic indicators to segment customers into targeted offer groups.

Assignment and Break

- Work through Part 1 of a visual introduction to machine learning
- http://www.r2d3.us/visual-intro-to-machine-learning-part-1/
- Break until 11:20

Problems Solved by Predictive Analytics

Applications of Predictive Analytics in Business

- Marketing
- Customer relations
- Underwriting
- Fraud detection
- Risk management
- Health care

Marketing

- Marketing Campaigns
- Product Placement
- Customer Behavior
- Customizing Content
- Associated Sales

Customer Relations

- Lower Churn (Non-Repeat Customers)
- Forecasting how many employees are needed
- Detecting faulty products
- Complaint tracking and routing

Banking

- Underwriting / Risk Management
- Fraud detection
- Customer acquisition and retention
- Collections

Healthcare

- Risk scoring for patient diseases
- Lowering readmissions
- Appointment no-shows
- Increasing patient data security

Around the Room

1. What are some examples of predictive models that have seen in the past in your everyday life?
2. What are some examples of predictive models that you have used at work?

Types of Problems Solved By Predictive Analytics

- Time Series Analysis (Forecasting)
  - Statistical technique that deals with time or trending data that is in a series of time periods or intervals
- Regression
  - Measuring of the relationship between the mean value of one variable (target) and corresponding values of other variables (features)
- Classification
  - Machine learning algorithms that learn how to assign a class label (target) from the corresponding variables (features)

Types of Problems Solved By Predictive Analytics

- Clustering
  - Given a set of attributes or a set of interactions - such as a provider referring a patient to another - creating mathematical models that group items by similarity or frequency of interaction.
  - Randomly guesses and then makes iterations to improve
  - Cluster Simulation (video)

- Association / Market Basket
  - Rule-based machine learning method for discovering interesting relationships between variables
  - Modelling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items.
  - Grocery Store

Types of Problems Solved By Predictive Analytics

- Image Recognition
  - Identifying places, people, objects within an image, and drawing conclusions from them by analyzing them.
  - Using prior images (data) to learn about the image or about future images.
  - Image classification has become one of the key pilot use-cases for demonstrating machine learning.
  - Now being used to recognize video
  - Cats vs. Dogs
  - Facebook Alternative Text

Types of Problems Solved By Predictive Analytics

- Natural Language Processing
  - Identifying words, phrases and meaning from within text or speech data
  - Given a set of text data, using software to automatically categorize and extract meaning from the text, such as sentiment or topics.
  - Now being used to analyze speech as well
  - Surveys, open ended text boxes
  - Do you think this handles structured or unstructured data?
  - London
  - Stitch Fix

Types of Problems Solved By Predictive Analytics

- Neural Networks
  - Algorithms that recognize underlying relationships in a set of data through a process that mimics the way the human brain operates
  - NN in 1 Minute (video)

- Deep Learning
  - Neural Networks have been around for decades, but were generally disappointing in terms of learning.
  - Improvements in mathematical formulas and increasingly powerful computers, we can now model many more layers of virtual neurons than ever before.

Types of Problems Solved By Predictive Analytics

- Large Language Models
  - ChatGPT
    - A generative, pre-trained transformer that uses NLP driven by AI
    - Allows users to to have conversations with a chatbot
    - Can create content like articles, summaries, essays, etc.

Around the Room

- How much have you learned about ChatGPT before today?
- Have you used ChatGPT before today?
- Can this be helpful for a data analyst or in predictive modeling?
    - How Ken Jee uses ChatGPT

Discussion Board and Lunch Time

- Google each of the 8 types of problems that are solved by predictive analytics, followed by your area of interest: (ex. Image Recognition in Education)
- You do not need to go into deep detail of the results. Just get a gist of what is happening in your field of interest
- Fill out the Current Topics in Predictive Analytics discussion board post with what you find. Reply to 1 other person.
- Meet back at 1:00 PM

Time-Series (Forecasting)

UWM
SCHOOL OF CONTINUING EDUCATION

Forecasting Model Framework

- **Do you need a predictive model?**
- Do you have a predictive model built?
- Does your predictive model perform as expected on relevant population?
- Do you have an implementation and workflow strategy?

Forecasting

- Answers the question, “How much or how many for the next time?”

- Looks at a time series data of numbers and predicts the future value for the data looking at that the trends.
  - Bayes Theorem – Probability of an event is based on prior knowledge
  - Predictive Analytics – Looking for trends and patterns in the past

Forecasting

- Difference is that forecasts should be very specific to what you are predicting and that you need as much data as possible.

- So, when we are predicting UW Health Turnover, the #1 feature is going to be past UW Health Turnover.

Forecasting

- You can create lags to determine when certain events trend up and down
- Autocorrelations and Partial-Autocorrelations can use the previous data along with the current data
- Predictions are then made for upcoming values based on previous values + seasonality + trends + randomness

Forecasting Example

- Number of HS graduates from past 25 years predicts number of HS graduates next year
- What happens when a new data set is used?
- What happens when another iteration is run?

Forecasting Model Framework

- Do you need a predictive model?
- **Do you have a predictive model built?**
- Does your predictive model perform as expected on relevant population?
- Do you have an implementation and workflow strategy?

Demo

Forecasting.ipynb

- Explain data
- Loading a csv
- Pandas Profiling
- Train/Test Splits
- Stationarity
- Seasonality
- Auto-Arima models
- Using Regressors
- Future Predictions

Demo

Pycaret Forecasting

- Sample Notebook Download
- Loading data
- Preparing data
- Setup
- Plots of data
- Model comparison
- Future Predictions

Break Time

- Meet back at 2:40 PM

Forecasting Model Framework

- Do you need a predictive model?
- Do you have a predictive model built?
- **Does your predictive model perform as expected on relevant population?**
- Do you have an implementation and workflow strategy?

Forecasting Evaluation Metrics

- Log Likelihood
  - The logarithm of the probability of the observed data coming from the estimated model. For given values of p, d and q
  - This is what is trying to be maximized when finding parameter estimates.

Forecasting Evaluation Metrics

- AIC
  - Estimate of a constant plus the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model
- BIC
  - Estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup
- AIC and BIC
  - Penalized-likelihood criteria
  - Lower value is closer to truth

Forecasting Evaluation Metrics

- Each feature in model (AR, MA, I, Regressors)
- Confidence Intervals
- Date of Prediction
  - You will have multiple predictions for a date
  - The predictions should get more accurate the closer to the date of prediction that you get

Forecasting Model Framework

- Do you need a predictive model?
- Do you have a predictive model built?
- Does your predictive model perform as expected on relevant population?
- **Do you have an implementation and workflow strategy?**

Around the Room

- Talk about examples you have seen of forecasting at work
- Talk about how the request was made to create a forecast model
- Talk about how the end-user or stakeholder ended up using the results

Places to find Data sets open to the public

- United States Government Open Data
- Kaggle
- Kapsarc
- ICPSUR

Assignment

- Create a new folder for holding your notebooks
- Copy the forecasting notebook into your notebooks folder
- Find a dataset that with Date, Target and at least one regressor
- Run your data through the notebook without using any regressors. Record the metrics.
- Run your data through the notebook with using regressors. Record the metrics.

Around the Room

- Brief overview of your dataset
- Talk about what model ended up being best without regressors
- Did the regressors help?
- What did you learn?

Summary of Day 1

UWM SCHOOL OF CONTINUING EDUCATION

Summary of Day 1

- Data Science Tools like Anaconda, R, Python, Notebooks, GitHub
- Descriptive, Diagnostic, Predictive and Prescriptive Analytics
- Predictive Model Framework
  - Do you need one?
  - Do you have one?
  - Does it work correctly?
  - Can you implement it?

Summary of Day 1

- Dependent vs Independent Variables
- Statistical Models vs Machine Learning
- Different types of Predictive Models (ex. Forecasting)
- Deep Dive into Forecasting

Homework