Project Workshop

July 2018

Final Project

- Thursday
- 5min presentation on your work (bell at 4min)
- Please submit a zip file via Slack with your name including:
  - Jupyter notebook for Final Project Part 1
  - Jupyter notebook for Final Project Part 2
  - Jupyter notebook for Final Project Part 3
  - Powerpoint/Keynote Final Project Part 4
- Audience will be class (with pizza and beer)
- Audience may ask 2 questions to presenter
- Invite employers, friends, family etc...

Final Project

Present your findings to an audience that may include non-technical executive stakeholders. Summarize and break down your problem, approach, and recommendations in a manner that different levels of expertise will find informative and persuasive.

Create a presentation that walks through your goals, approach, results, and recommendations. Keep slides simple - no more than 2-3 pieces of information per slide. When in doubt, use visuals!

Finally, be prepared to explain and defend your model to an inquisitive audience.
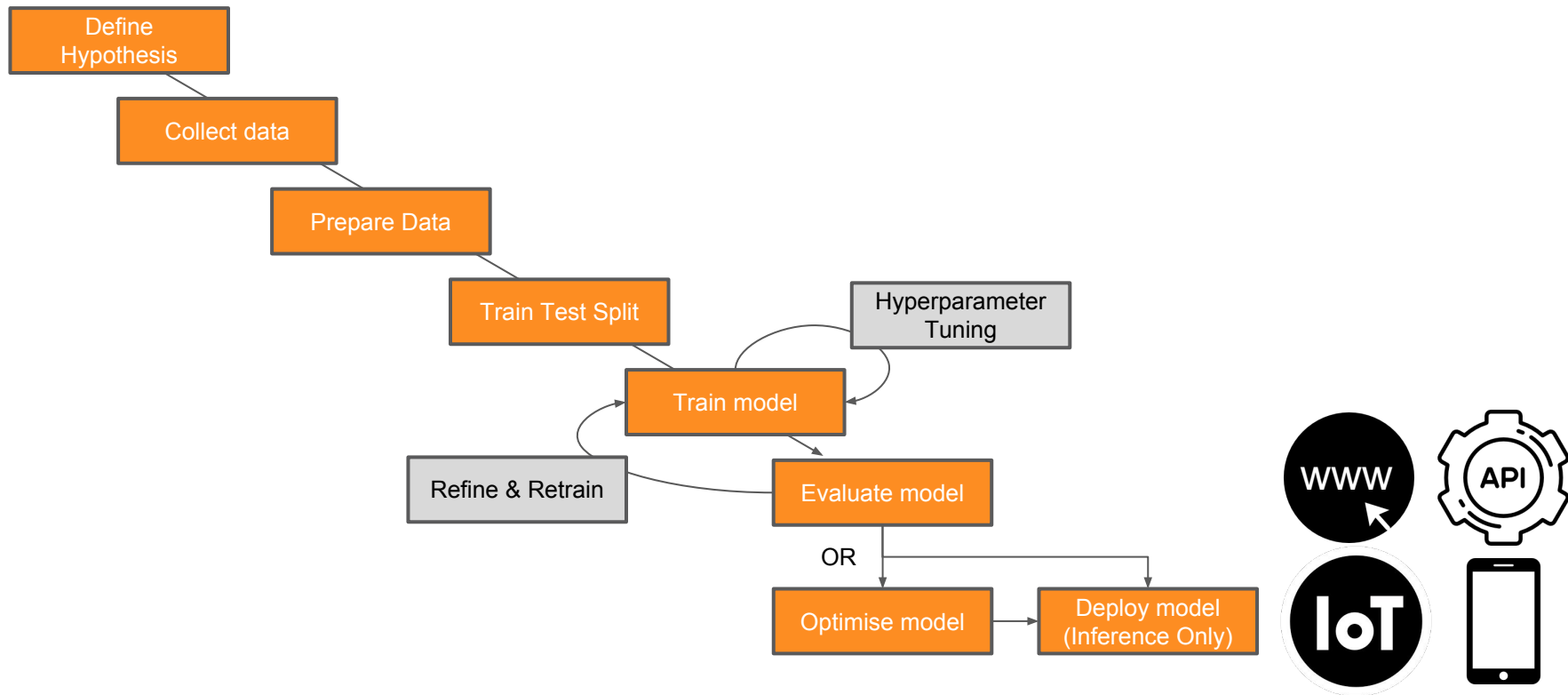
Final Project

Requirements

1. Detailed presentation for non-technical audience.
2. Your presentation should include:
   - A problem statement.
   - Metrics and assumptions.
   - Approach and process.
   - Your model or solution.
   - Performance evaluation (how your approach compares to your original success metrics).
   - Impact of your findings.
   - Recommendations or next steps.

ML Process

Model Building Steps

1. Create hypothesis
2. Locate and access data
3. Subset to appropriate time period, e.g. number of stores or a 10% sample
4. Explore data e.g graphs and correlations
5. Clean data (fix variable types e.g dates etc, parse strings, create dummy variables)
6. Normalise and scale
7. Test train split
8. Check distribution of positive + negative cases in train and test sets to make sure it's balanced
9. Select model features
10. Kfolds
11. Build models (Linear Regression, Logistic Regression, RandomForestClassifier,)
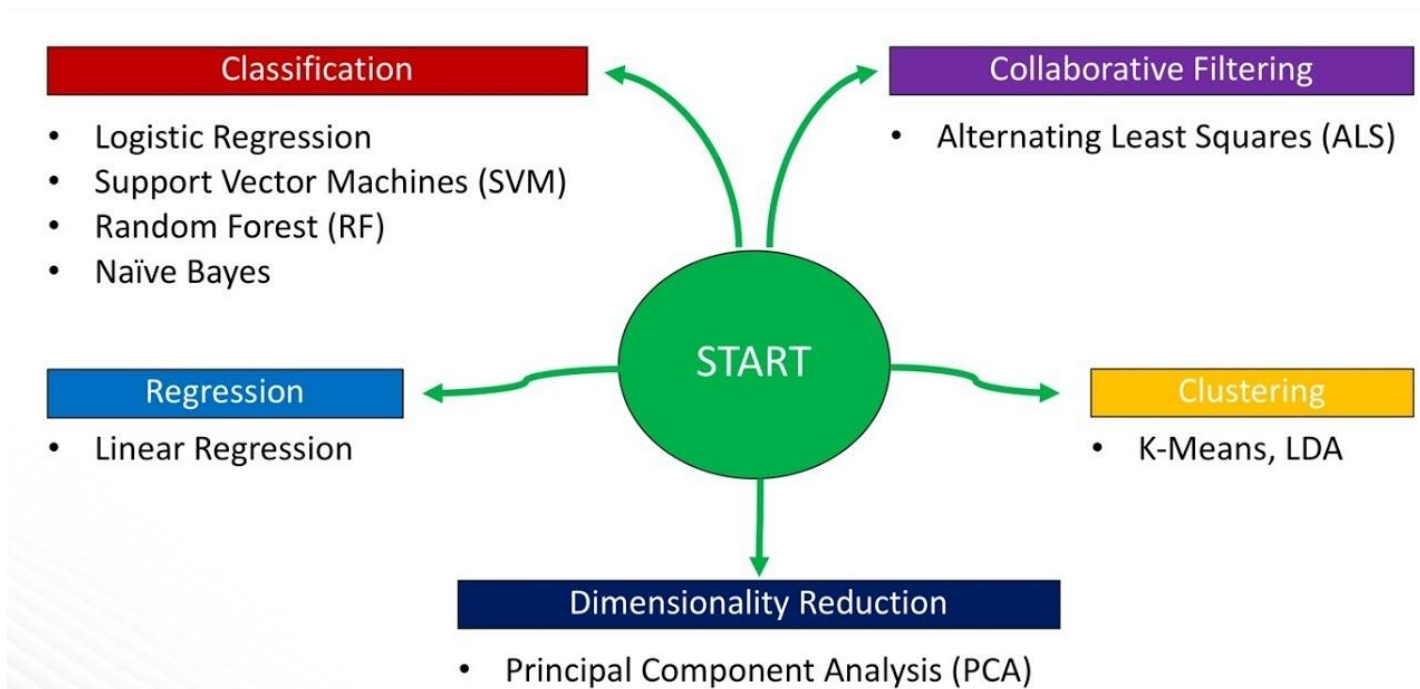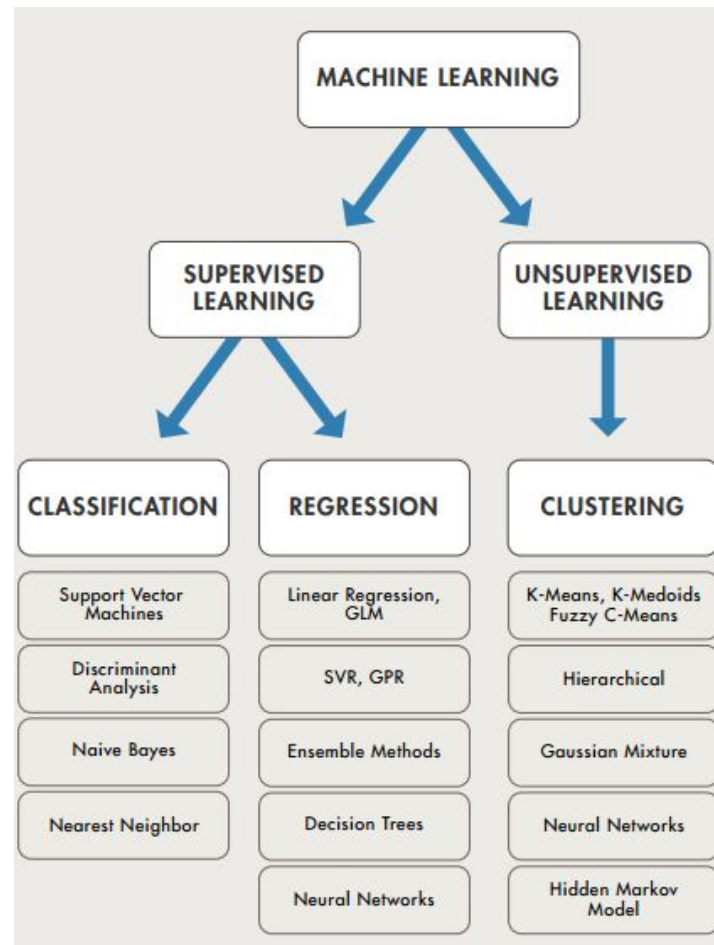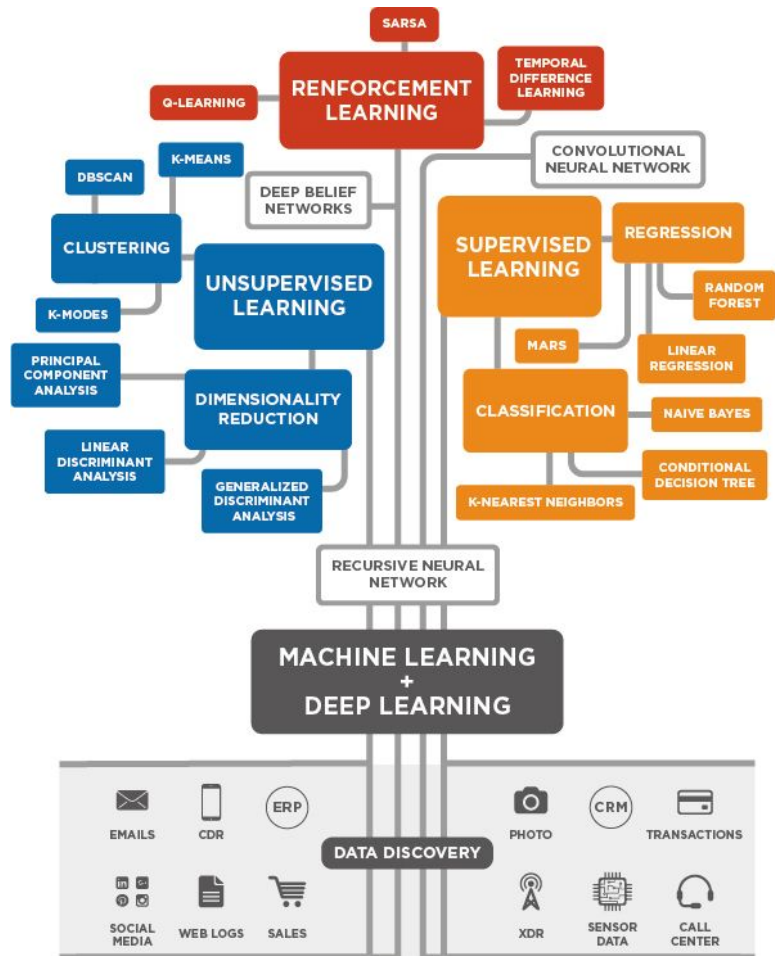12. Evaluate models

Data Cleaning

Data Cleaning

- Numbers to int or float
- Remove '$', '%', '+', '-'
- Convert strings to dates
- Make new columns/features for Year, Month, Day-of-Week
- Days since event e.g. Days since last purchase = Today() - Last_purchase_date
- Tenure e.g. Today() - Start_Date

Model Zoo

Model Zoo

Left diagram:

SARSA

RENFORCEMENT LEARNING

Q-LEARNING

TEMPORAL DIFFERENCE LEARNING

CONVOLUTIONAL NEURAL NETWORK

K-MEANS

DBSCAN

DEEP BELIEF NETWORKS

CLUSTERING

UNSUPERVISED LEARNING

SUPERVISED LEARNING

REGRESSION

RANDOM FOREST

K-MODES

MARS

LINEAR REGRESSION

PRINCIPAL COMPONENT ANALYSIS

DIMENSIONALITY REDUCTION

CLASSIFICATION

NAIVE BAYES

LINEAR DISCRIMINANT ANALYSIS

GENERALIZED DISCRIMINANT ANALYSIS

K-NEAREST NEIGHBORS

CONDITIONAL DECISION TREE

RECURSIVE NEURAL NETWORK

MACHINE LEARNING + DEEP LEARNING

EMAILS CDR ERP PHOTO CRM TRANSACTIONS

SOCIAL MEDIA WEB LOGS SALES XDR SENSOR DATA CALL CENTER

DATA DISCOVERY

Right diagram:

MACHINE LEARNING

SUPERVISED LEARNING

UNSUPERVISED LEARNING

CLASSIFICATION	REGRESSION	CLUSTERING
Support Vector Machines	Linear Regression, GLM	K-Means, K-Medoids Fuzzy C-Means
Discriminant Analysis	SVR, GPR	Hierarchical
Naive Bayes	Ensemble Methods	Gaussian Mixture
Nearest Neighbor	Decision Trees	Neural Networks
	Neural Networks	Hidden Markov Model

Random Forest

Random Forest Notebook

Markdown ▾   📊 Enter/Exit RISE Slideshow

Introduction to Decision Tree and Random Forest Intuition

Outcomes

- Explain how a decision tree functions and how to build one
- Visually represent a decision tree
- Determine when using a decision tree is an appropriate method

Let's look at an example

Will Jonny play tennis on a particular day? Give the attributes of weather outlook, temperature, humidity, and wind strength Jonny makes a decision whether he will play tennis or not. Can we predict his decision?

```
In [77]: train = pd.read_csv('data/play_tennis.csv')
         info = pd.read_csv('data/play_tennis_info.csv')
```

```
In [78]: info
```

Out[78]:

	Attribute	Possible Values
0	Outlook	sunny, overcast, rain
1	Temperature	hot, mild, cool
2	Humidity	high, normal
3	Windy	true, false

Q&A

EXIT TICKETS

DON'T FORGET TO FILL OUT YOUR EXIT TICKET

Learning Objectives