Communication

- Email: ysalomon@(ischool.)berkeley.edu (when in need - [URGENT]...)
- Slack: ucbischool.slack.com #w207-2018-0904-s1
- LMS: learn.datascience.berkeley.edu
  - Wall
  - In class chat public/private
  - Files -- Google docs
- Github: ysalomon/ucb-w207-applied-ml

Projects and due dates

3 problem sets

- Problem set 1: due date **October 1st, 2018**
- Problem set 2: due date **October 29th, 2018**
- Problem set 3: due date **November 26th, 2018**

Code @ https://github.com/ysalomon/ucb-w207-applied-ml/tree/master/projects

Final project

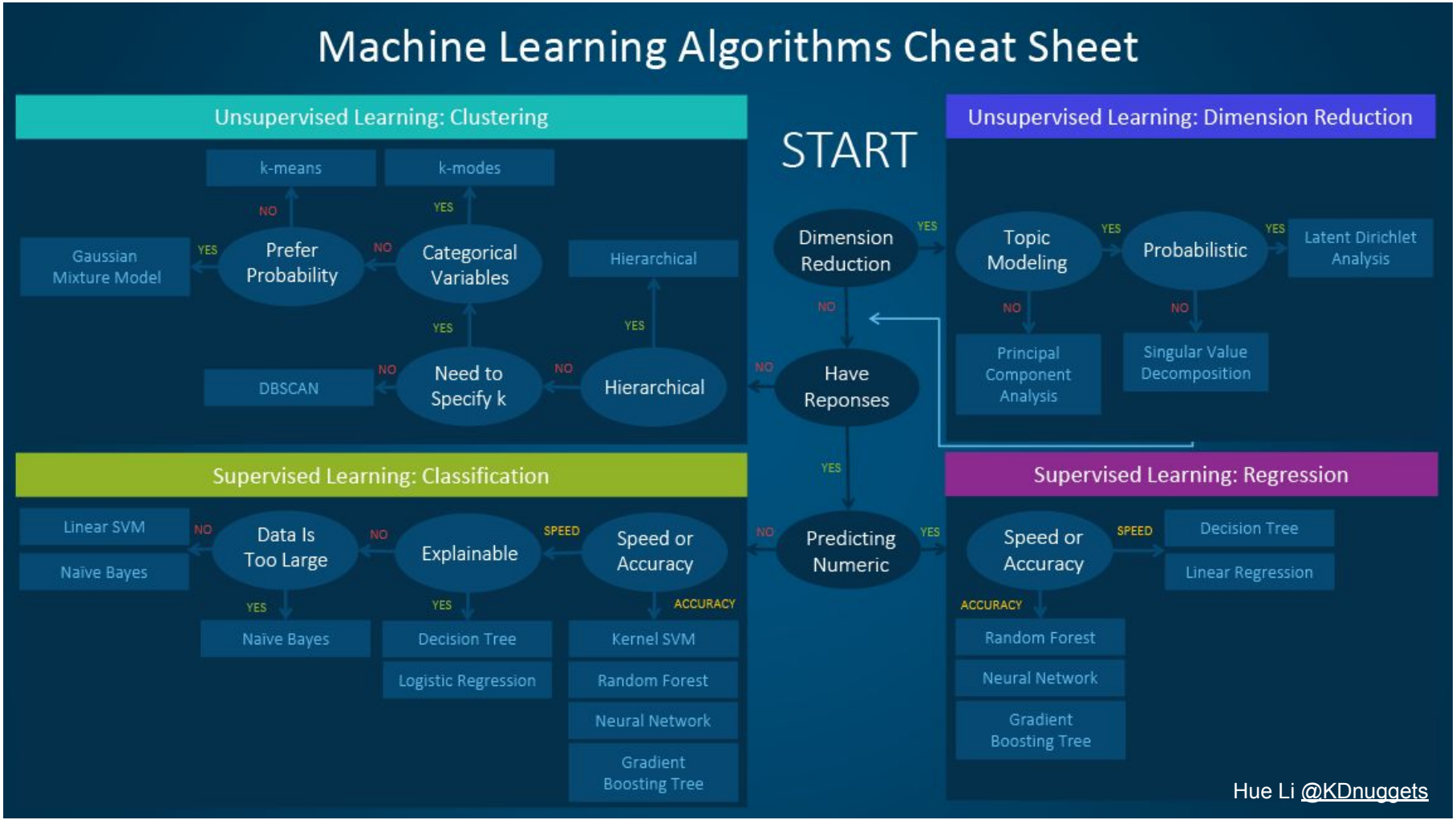
- Details will be shared on **October 15th, 2018**
- Kaggle/datasets

Learning goals

- Theory
  - Terminology
  - Demystifying - overarching principles
  - Underlying mathematics
- Application
  - Familiarity with the tools
  - Workflow
  - In the real-world ...
- Habits
  - How to approach a problem
  - Where to find information

Course overview

- What is ML / AI / data-science ? Supervised vs Unsupervised ?
- Supervised learning algorithms:
  a. kNN
  b. Naive Bayes
  c. Decision trees
  d. Regression and regularization
  e. Optimization with gradient descent
  f. Neural networks and backpropagation
- Unsupervised learning algorithms:
  a. Clustering methods - k-means & hierarchical clustering
  b. EM algorithms and mixture models (topic models/LDA)
  c. Social network analysis


Machine Learning Algorithms Cheat Sheet
START
Unsupervised Learning: Clustering
k-means
k-modes
NO
YES
Gaussian Mixture Model
YES
Prefer Probability
NO
Categorical Variables
Hierarchical
YES
YES
DBSCAN
NO
Need to Specify k
NO
Hierarchical
Dimension Reduction
YES
NO
Have Reponses
NO
YES
Unsupervised Learning: Dimension Reduction
Topic Modeling
YES
Probabilistic
YES
Latent Dirichlet Analysis
NO
NO
Principal Component Analysis
Singular Value Decomposition
Supervised Learning: Classification
Linear SVM
Naïve Bayes
NO
Data Is Too Large
NO
Explainable
SPEED
Speed or Accuracy
YES
YES
ACCURACY
Naïve Bayes
Decision Tree
Logistic Regression
Kernel SVM
Random Forest
Neural Network
Gradient Boosting Tree
NO
Predicting Numeric
YES
Supervised Learning: Regression
Speed or Accuracy
SPEED
Decision Tree
Linear Regression
ACCURACY
Random Forest
Neural Network
Gradient Boosting Tree
Hue Li @KDnuggets