Week 11: Model Based Clustering
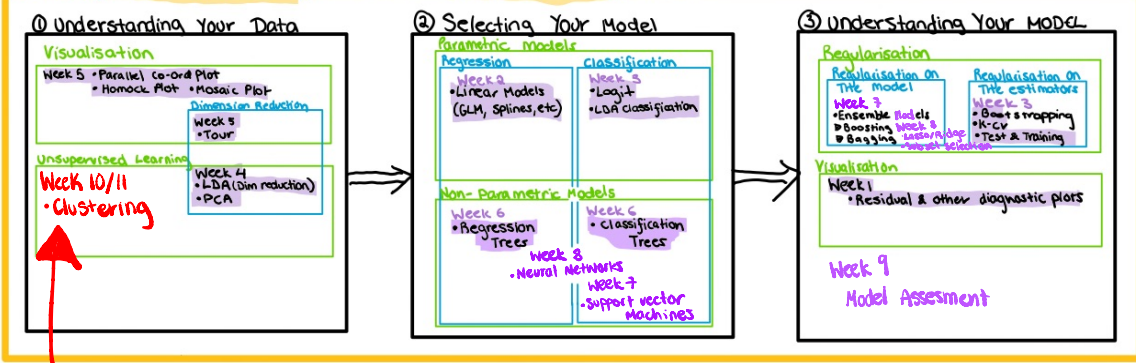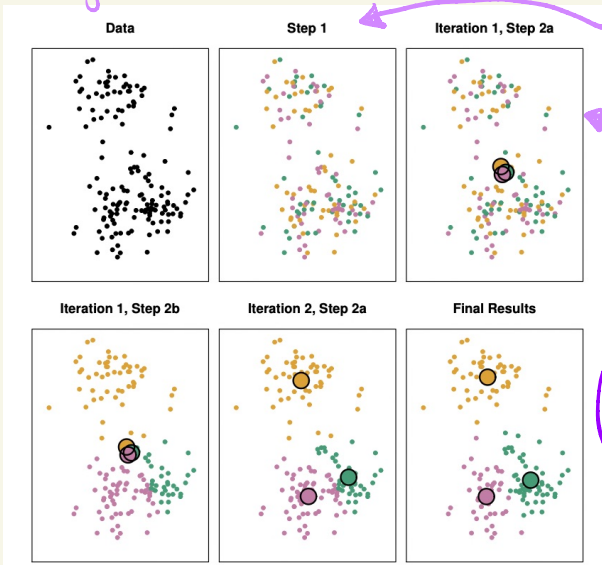
① Big Picture

◎ Things that influence your Bias and Variance trade off (week 1)

① Understanding Your Data

Visualisation
- Week 5 • Parallel Co-ord Plot
  • Homock Plot • Mosaic Plot

Dimension Reduction
- Week 5 • Tour

Unsupervised Learning
- Week 10/11
  • Clustering

- Week 4
  • LDA (Dim reduction)
  • PCA

② Selecting Your Model

Parametric Models

Regression
- Week 2
  • Linear Models (GLM, Splines, etc)

Classification
- Week 5
  • Logit
  • LDA classification

Non-Parametric Models
- Week 6
  • Regression Trees
- Week 6
  • Classification Trees

Week 8
• Neural Networks
Week 7
• Support Vector Machines

③ Understanding Your Model

Regularisation

Regularisation on The model
- Week 7
  • Ensemble Models
  ▷ Boosting Week 8 Lasso/ridge
  ▷ Bagging variable selection

Regularisation on The estimators
- Week 3
  • Boot strapping
  • K-CV
  • Test & Training

Visualisation
- Week 1
  • Residual & other diagnostic plots

Week 9
Model Assessment

→ We Are Here

② K-means Clustering
Algorithm

1. Randomly assign each Observation to a Cluster (1 to k)

2. Compute the Cluster Centroid

3. Assign each Observation to the Cluster whose Centroid is the Closest. (euclidean distance)

Repeat until Cluster assignments stop changing

Why set a seed?
The original random assignment Can impact the final result