Final Study Guide

UL2 - Clustering

- Single Linkage Clustering - what would happen if we chose furthest or average distance, rather than closest intercluster distance?
    - Would change the algorithm from “Single Linkage” to another type
- What is the running time for single linkage clustering?
    - O(n^3)
- What type of optimization algorithm does K-Means follow?
    - Hill-climbing
- Will K-Means always converge to the optimal clustering?
    - No, can get stuck in local optima
- Properties of EM algorithm?
    - Monotonically non-decreasing likelihood
    - Does not converge (although practically does)
    - Will not diverge
    - Can get stuck
    - Works with any distribution
- Clustering Properties (general, not just K-Means, EM)
    - Richness - for any assignment of objects to clusters, there is some distance matrix that would cause the clustering algorithm that would produce that clustering
    - Scale-invariance - if doubling / halving the distances, the clustering shouldn’t change
    - Consistency - if a clustering algorithm produces some algorithm: shrinking the intra-clustering distances and expanding the inter-clustering distances should not change the clustering
- Is there a clustering algorithm that can achieve all of three properties above?
    - No! Proven in Impossibility Theorem by Kleinberg
- How long does each clustering algorithm take?
    - SLC terminates in polynomial time
    - K-Means, EM can take longer
- Compare and contrast each of the 3 major clustering algorithms: SLC, K-Mean,s and EM
    - XXX

UL 3 - Feature Selection

- How long does it take (time complexity) to select m features from n features (where m <= n)?
    - Exponential - there are 2^n total possible subsets (NP-hard)
- Approach to Feature Selection: Filtering vs Search Algorithms
    - Filtering
        - + speed
        - - speed -> isolated features
        - - ignores the learning problem

    - Searching
        - + takes into account model biases
        - + …and learning method
        - - slow
- What is the smallest set of features sufficient to get zero training error?
    - See lectures for how to do
    - In general, start with one feature and add more features until you're able to get to the output
- Define relevance vs usefulness
    - Relevance measures effect on Bayes Optimal Classifier
    - Usefulness measures effect on a particular model or learning algorithm

UL 4 - Feature Transformation

- Feature Selection vs Feature Transformation?
    - Feature selection - selecting existing features
    - Feature transformation - create new feature set using any type of combinations (we restrict to linear transformations)
    - Feature selection is a subset of feature transformation
- PCA vs ICA quiz
    - Mutually orthogonal
        - PCA
    - Mutually independent
        - ICA
    - Maximal variance
        - PCA
    - Maximal mutual information
        - ICA
    - Ordered features
        - PCA
    - Bag of features
        - PCA, ICA
- Advantages of RCA?
    - Cheap (computationally), simple, easy, and FAST

RL 3 - Game Theory

- How many different strategies are there in a 2-player zero sum finite deterministic game of perfect information?
    - (# of options at each node) ^ (# of nodes where a choice is possible)
- Difference between pure strategy and mixed strategy?
    - Mixed strategy has a distribution over the strategies