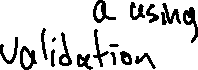
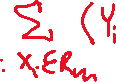
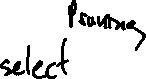
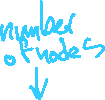
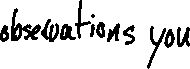
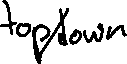
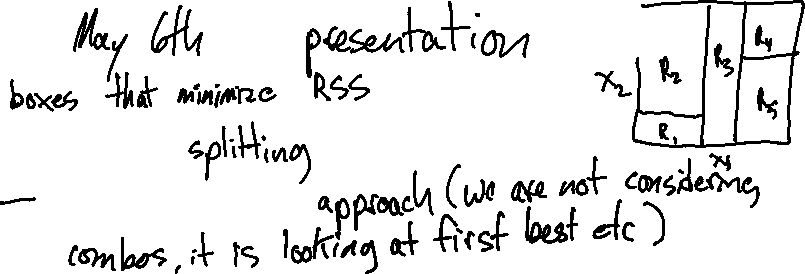
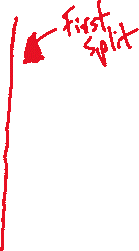
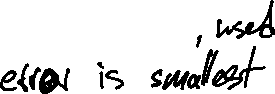
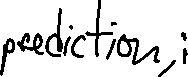
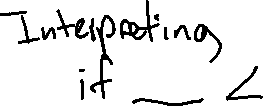


**Can trees account for values that are super rare?**

* It could immediately separate it self
* It also depends on how complex you let your tree be

**How can you correct for too-much complexity in a tree?**

* Play with the penalty alpha
* Cut based on number of nodes (more intuitive)



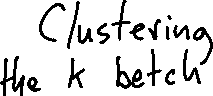
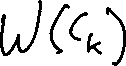
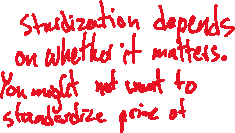
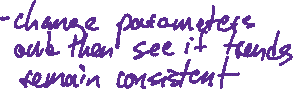
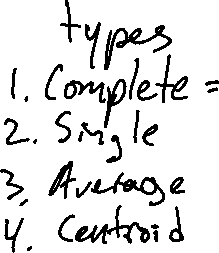
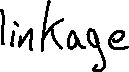
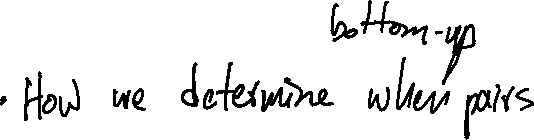
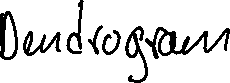
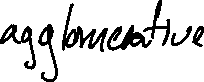
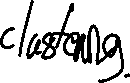
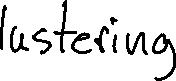
1. What is a distance matrix and what role do they play in hierarchical clustering?



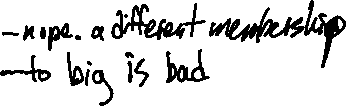
1. How can you be sure you have used the best unsupervised algorithm?



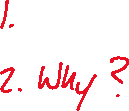
1. What information need to be conveyed to be convinced that clustering has done a good job.



1. Should you expect to find the same cluster memberships each time you do kmeans clustinng on the same data? What insigh do you have into how many clusters to use?



1. Discuss the similarities and differences between kmeans clustering and PCA. How do they compare on interpretability?



1. Discuss the role of standardization/scaling have on clusters? On PCA?

