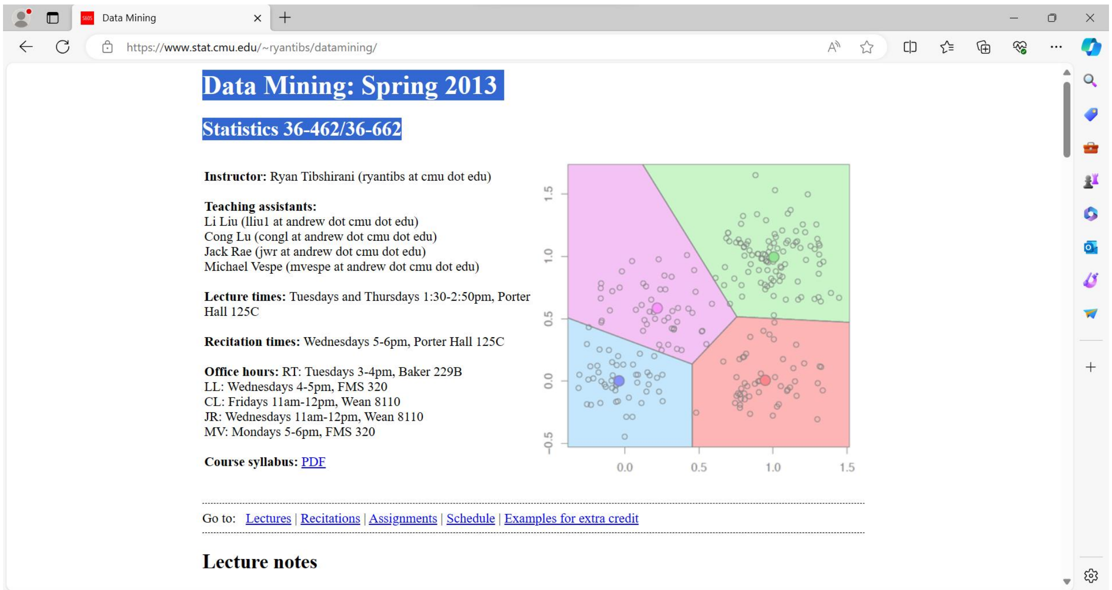
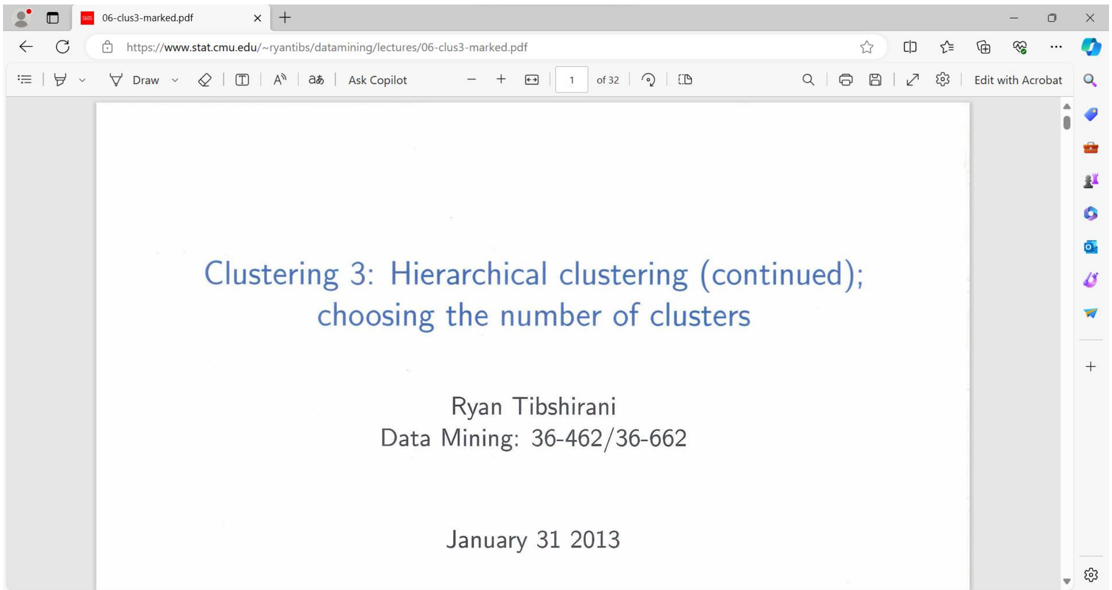
Carnegie Mellon University

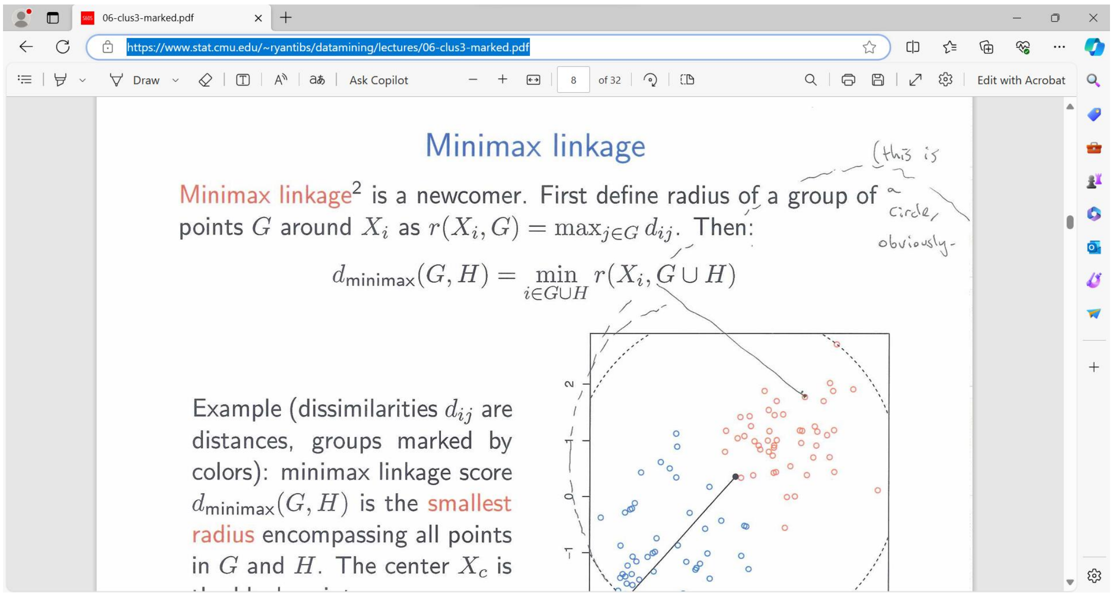
University Undergraduate Course: Data Mining

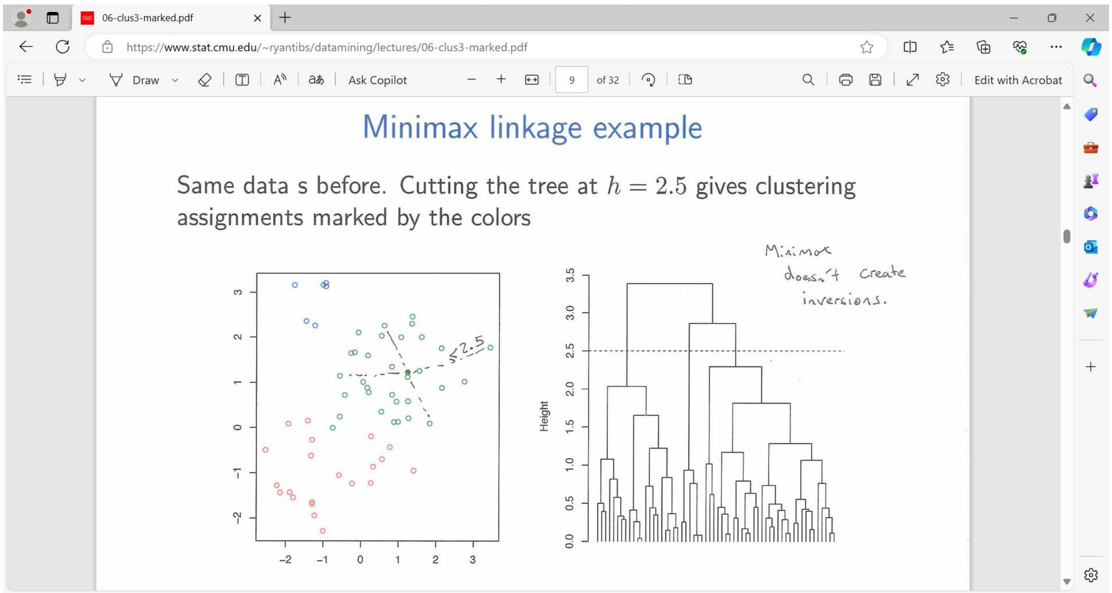
https://www.stat.cmu.edu/~ryantibs/datamining/
(retrieved 11 March 2024)



https://www.stat.cmu.edu/~ryantibs/datamining/lectures/06-clus3-marked.pdf
(retrieved 11 March 2024)



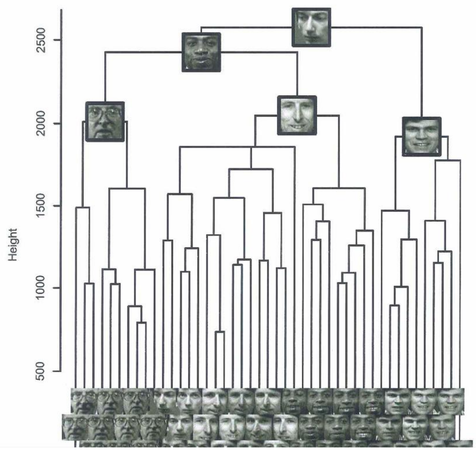

06-clus3-marked.pdf
https://www.stat.cmu.edu/~ryantibs/datamining/lectures/06-clus3-marked.pdf
Draw
Ask Copilot
8
of 32
Edit with Acrobat
Minimax linkage
Minimax linkage² is a newcomer. First define radius of a group of points G around X_i as r(X_i, G) = max_{j∈G} d_ij. Then:
d_minimax(G, H) = min_{i∈G∪H} r(X_i, G ∪ H)
(this is a circle, obviously.)
Example (dissimilarities d_ij are distances, groups marked by colors): minimax linkage score d_minimax(G, H) is the smallest radius encompassing all points in G and H. The center X_c is

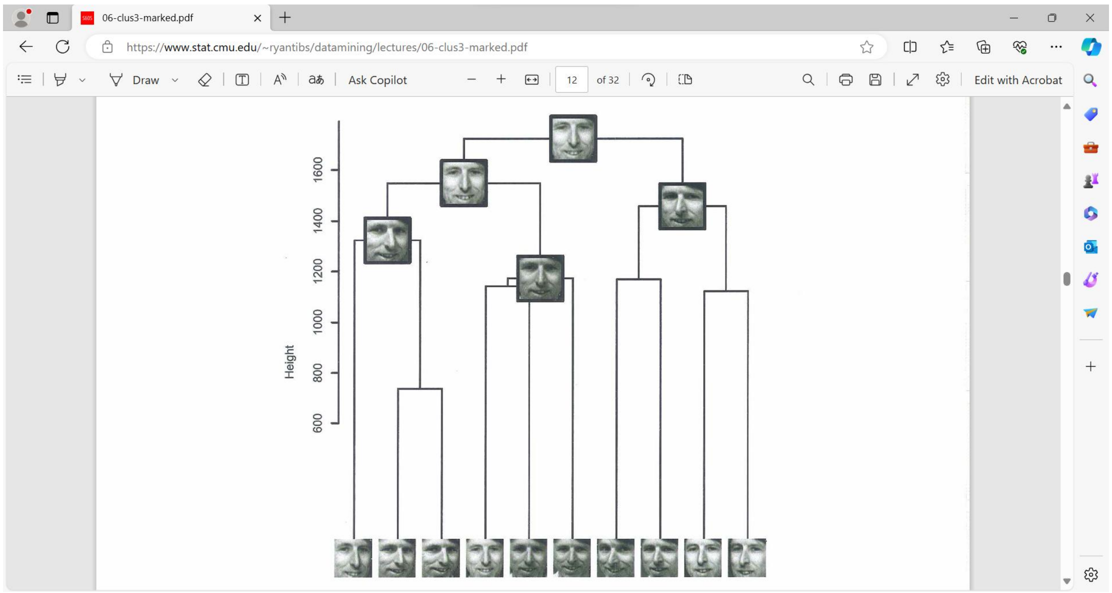

06-clus3-marked.pdf
https://www.stat.cmu.edu/~ryantibs/datamining/lectures/06-clus3-marked.pdf
Draw
Ask Copilot
9
of 32
Edit with Acrobat
Minimax linkage example
Same data s before. Cutting the tree at h = 2.5 gives clustering assignments marked by the colors
Minimax doesn't create inversions.
≤2.5
Height

Properties of minimax linkage

- Cutting a minimax tree at a height h a nice interpretation: each point is $\leq h$ in dissimilarity to the center of its cluster. (This is related to a famous set cover problem)
- Produces dendograms with no inversions
- Unchanged by monotone transformation of dissimilarities d_{ij}
- Produces clusters whose centers are chosen among the data points themselves. Remember that, depending on the application, this can be a very important property. (Hence minimax clustering is the analogy to K-medoids in the world of hierarchical clustering)

Example: Olivetti faces dataset





06-clus3-marked.pdf
https://www.stat.cmu.edu/~ryantibs/datamining/lectures/06-clus3-marked.pdf
Draw Ask Copilot 13 of 32 Edit with Acrobat

Centroid and minimax linkage in R

The function `hclust` in the base package performs hierarchical agglomerative clustering with centroid linkage (as well as many other linkages)

E.g.,

```
d = dist(x)
tree.cent = hclust(d, method="centroid")
plot(tree.cent)
```

The function `protoclust` in the package `protoclust` implements hierarchical agglomerative clustering with minimax linkage

Linkages summary

Linkage	No inversions?	Unchanged with monotone transformation?	Cut interpretation?	Notes
Single	✓	✓	✓	chaining
Complete	✓	✓	✓	crowding
Average	✓	×	×	
Centroid	×	×	×	simple
Minimax	✓	✓	✓	centers are data points

Note: this doesn't tell us what "best linkage" is

What's missing here: a detailed empirical comparison of how they