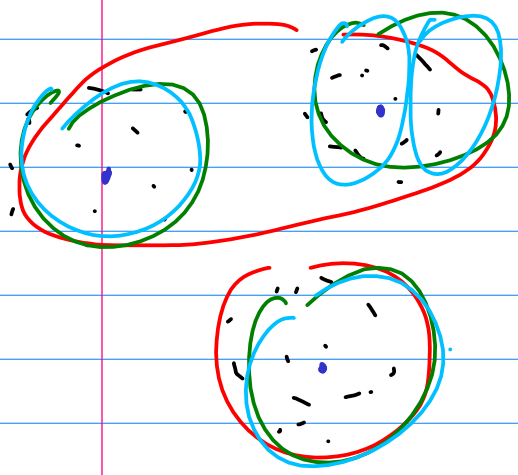
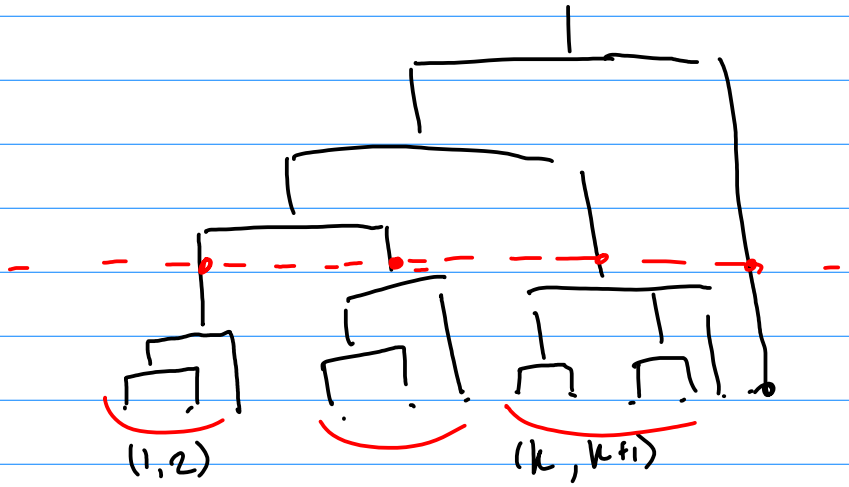
Clustering Data

K - means

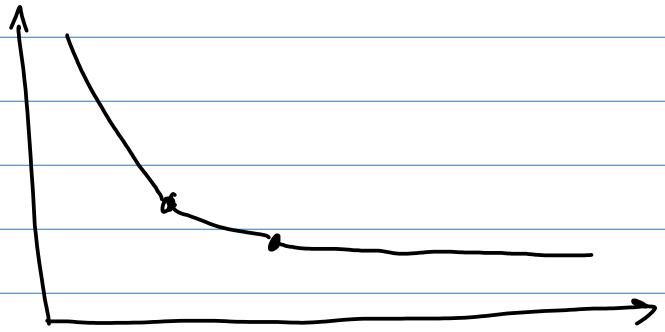


Hierarchical clustering



Objective function F

k	1	2	3	4	5	.....
F(k)	300	200	100	90	80	

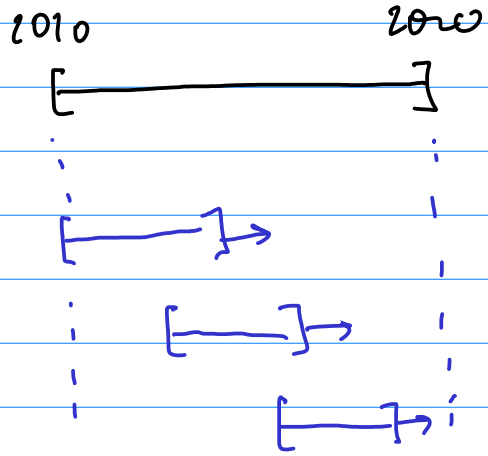


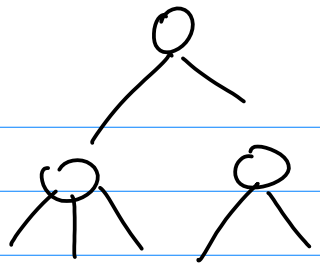
Tasks

1) Why? Multinomial Logit

2) Forecasting after clustering

3) Hierarchical Data





Country = Σ Regions

Regions = Σ Cities

Cities

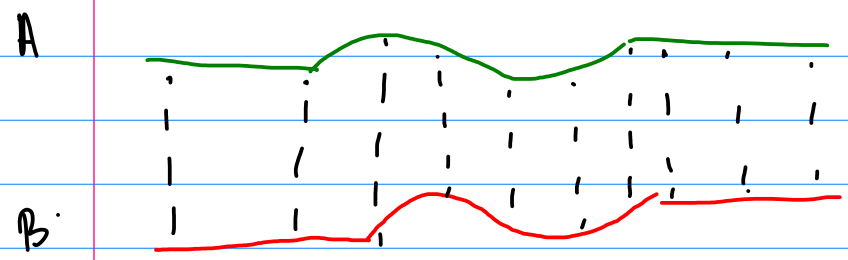
Forecasting : 1) Individually

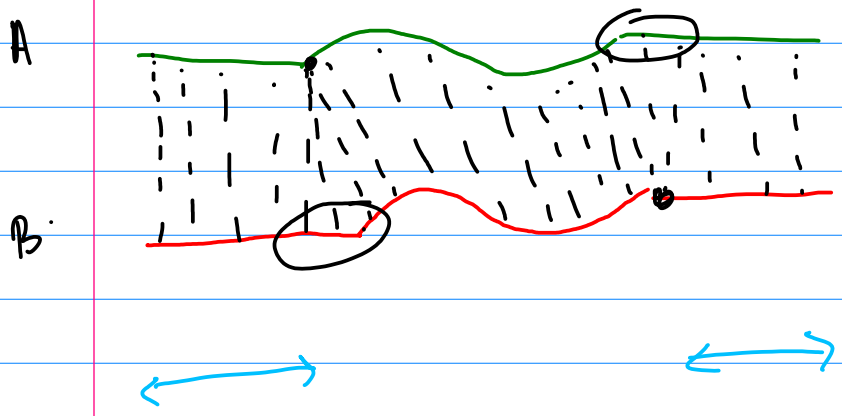
no

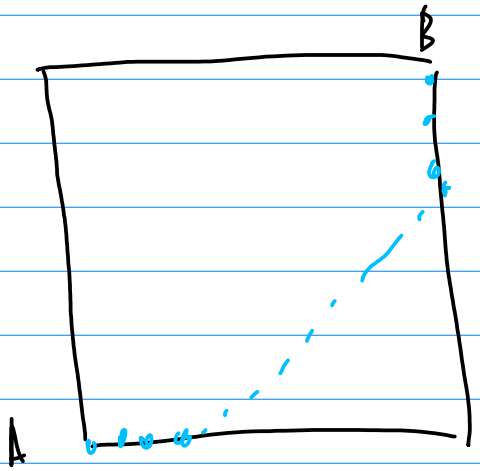
=> hierarchical structure

Distance Metric

1) Euclidian Distance
2) Correlation Metric
3) DTW (Dynamic Time Warp)







Distance Between Clusters

1) Single link (min distance)

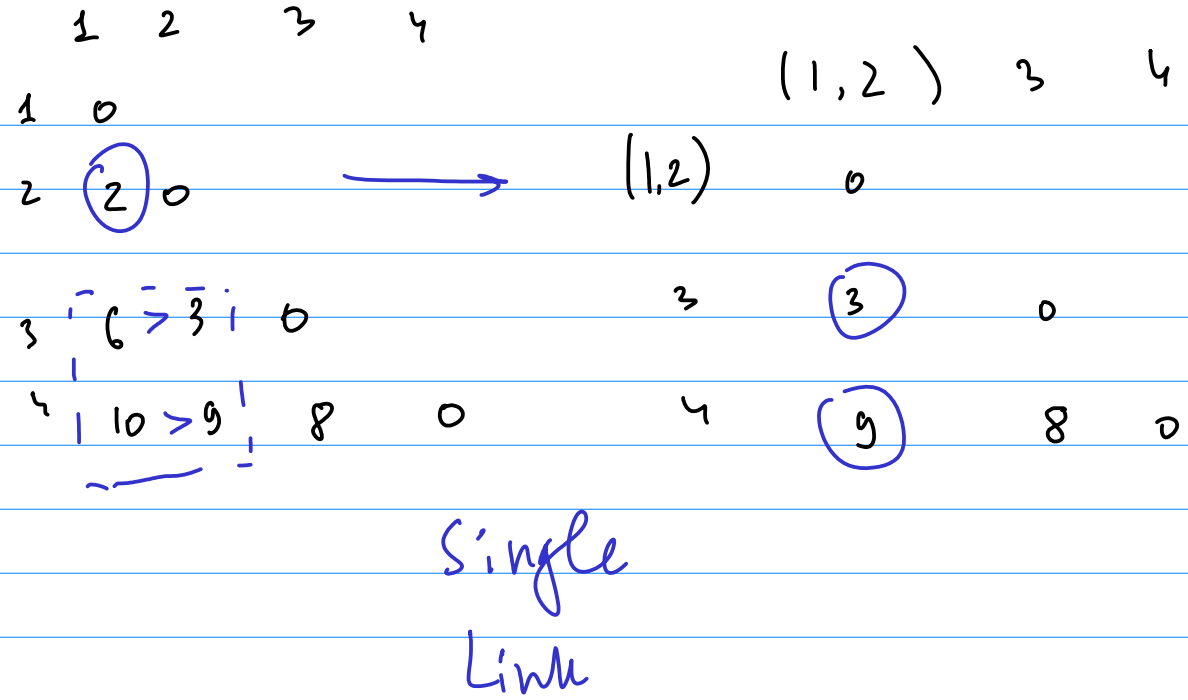
- long clusters

2) Complete link (max distance)

+ tight cluster

- not robust to outliers

3) Average Link (mean distance between all pairs)



Single Link