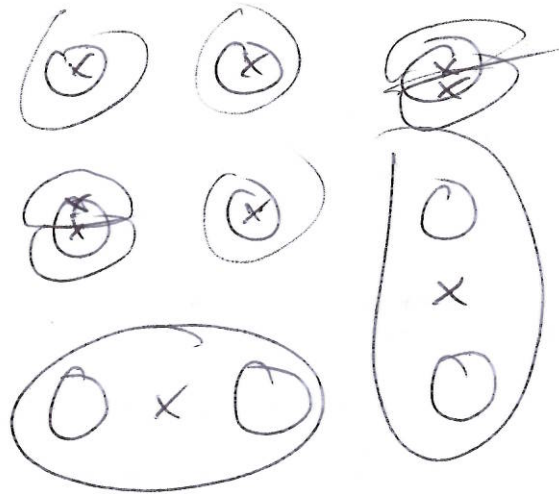


Kmeans + +

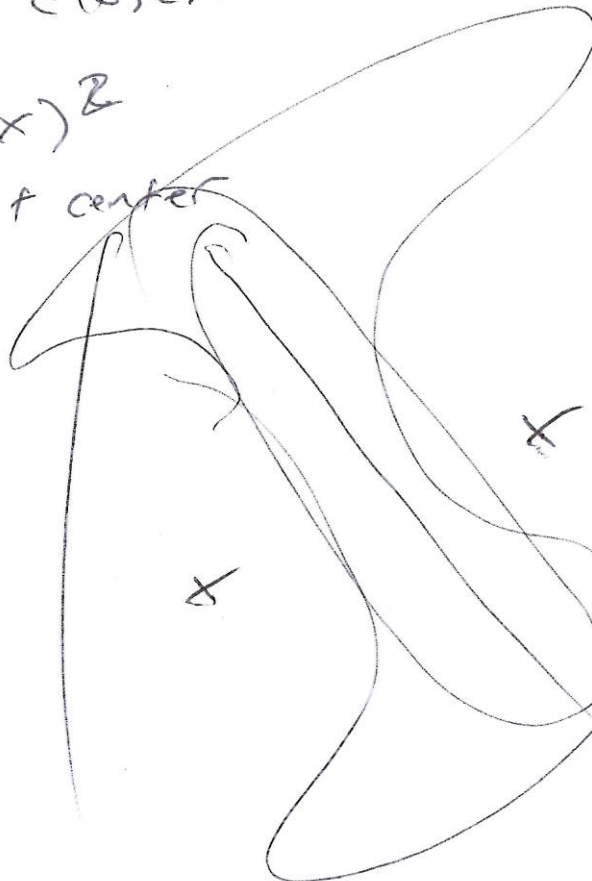


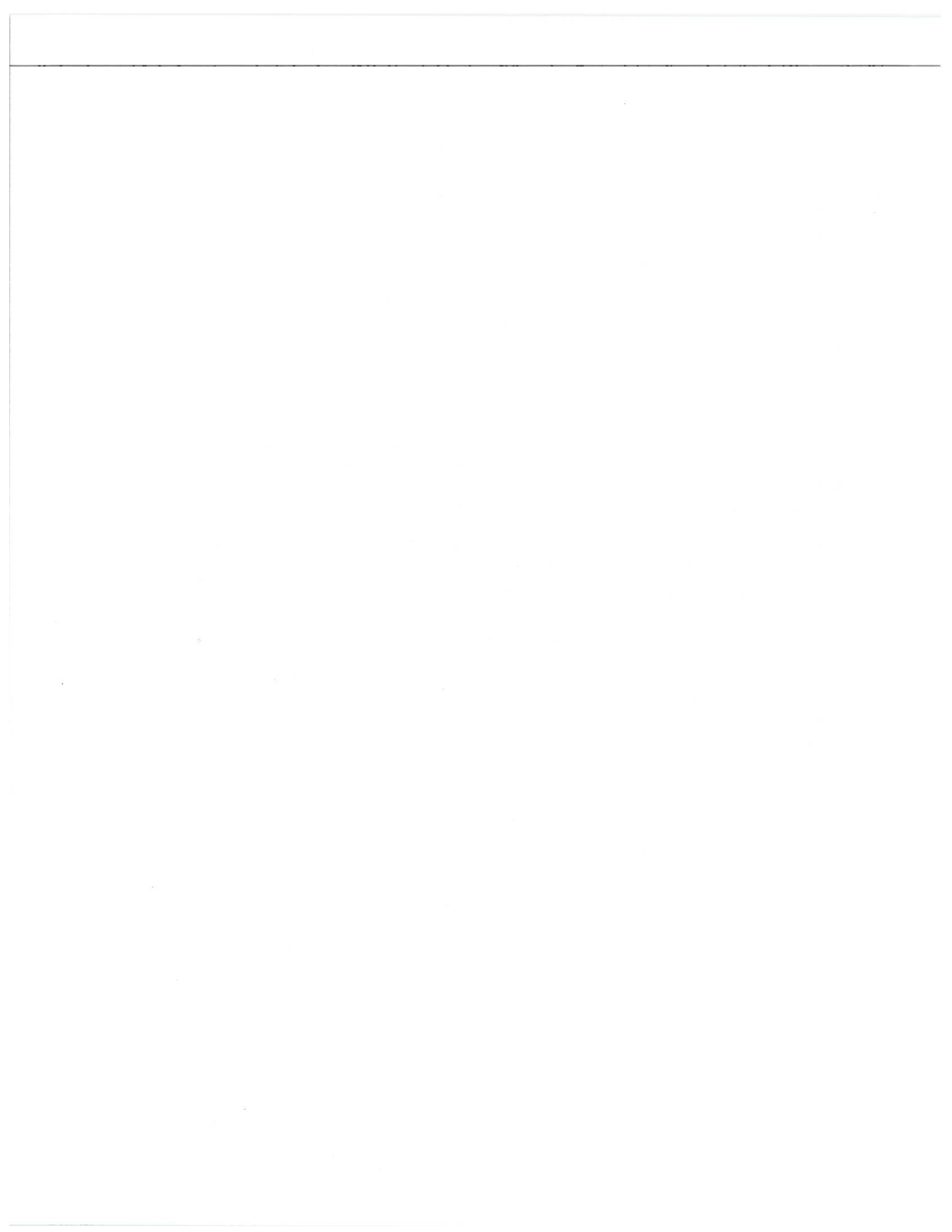
Pick first center

Compute $D(x)$ to closest center

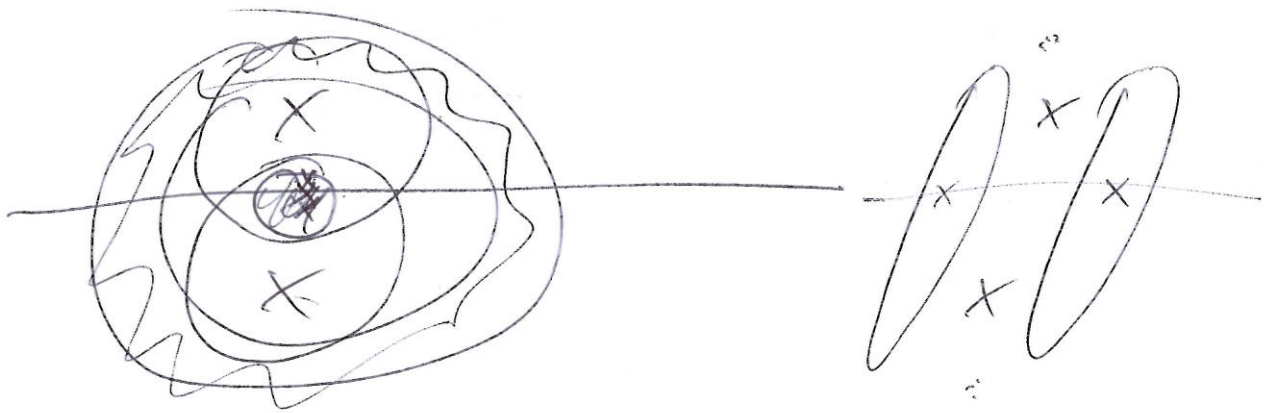
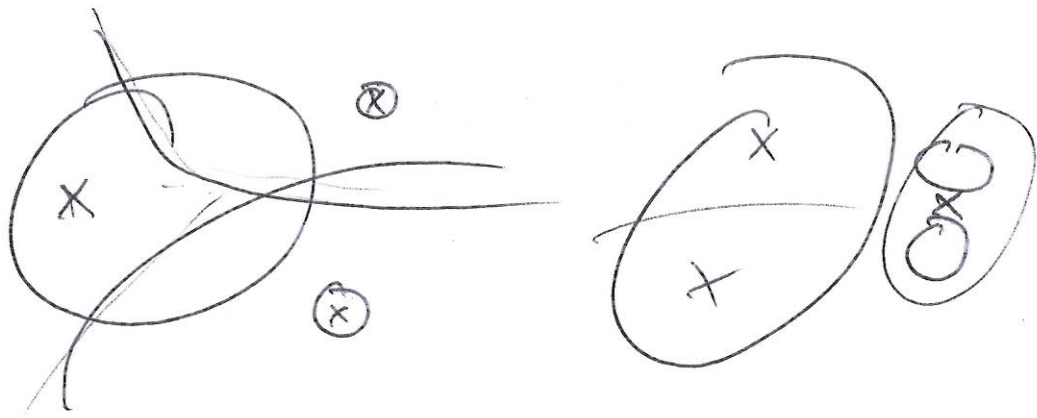
Set $P(x) \propto D(x)^2$

probability of picking x as next center





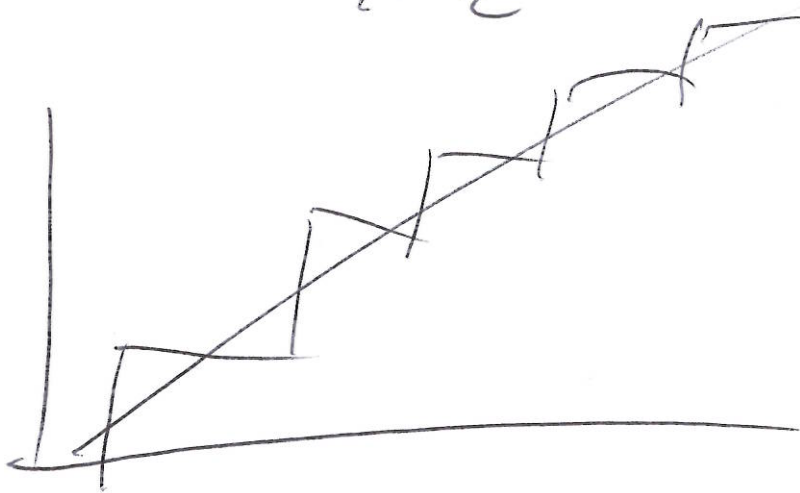
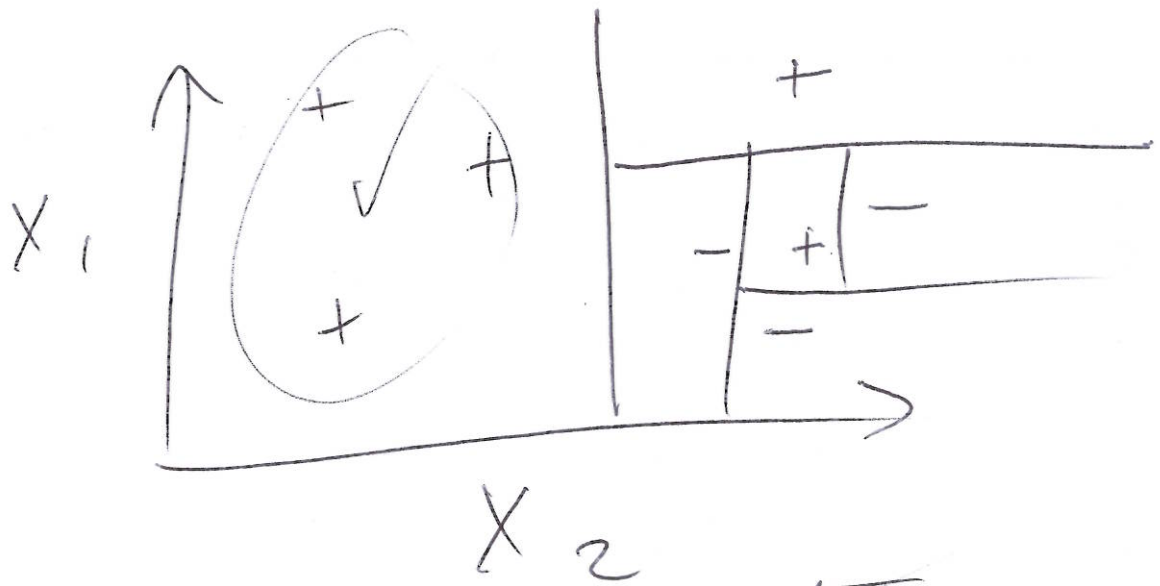
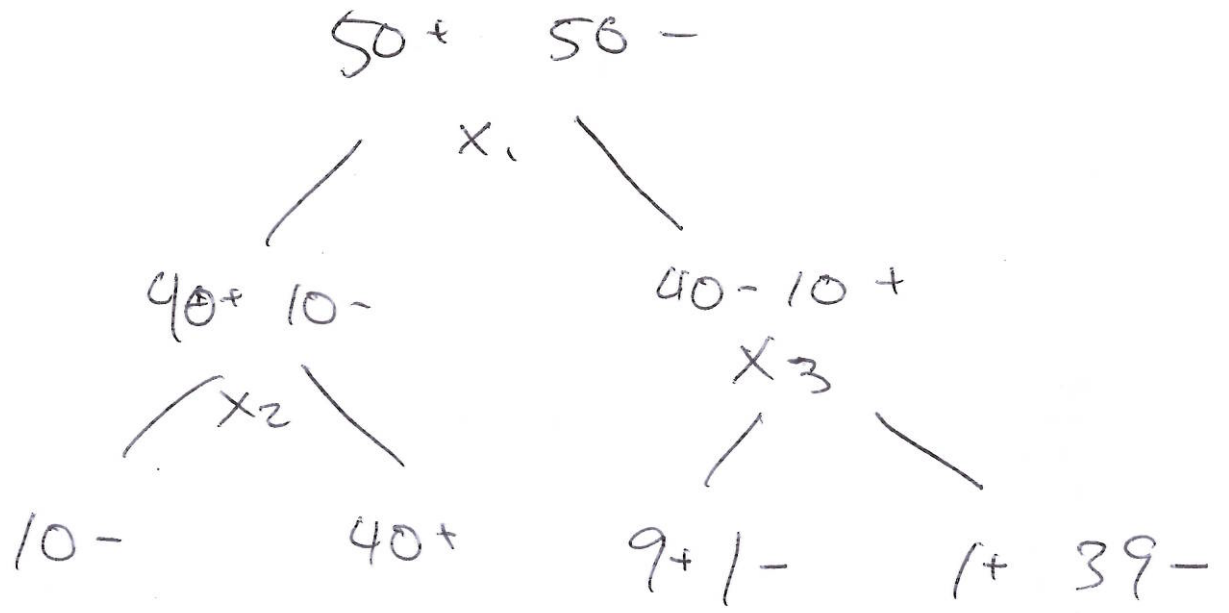




K-medoids

take median instead of mean





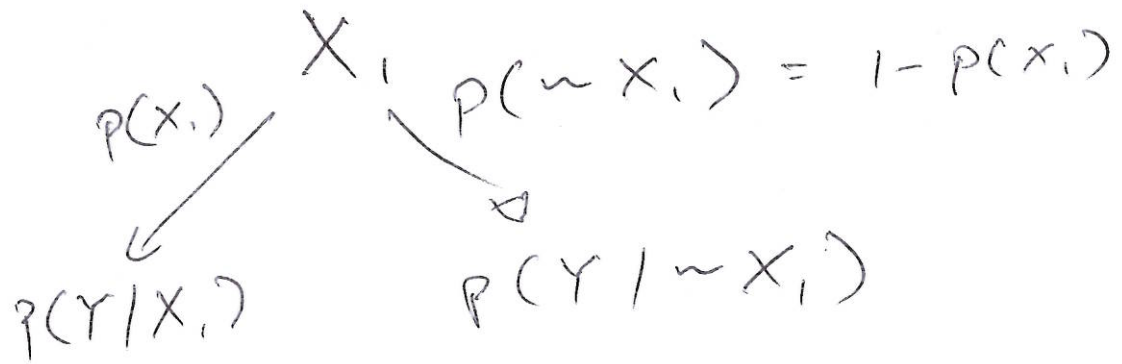
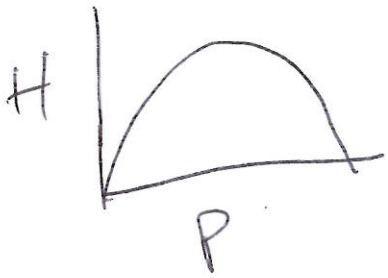
For each node

Record cond entropy
for each feature

- Pick minimum
- Split tree on that feature

Binary entropy:

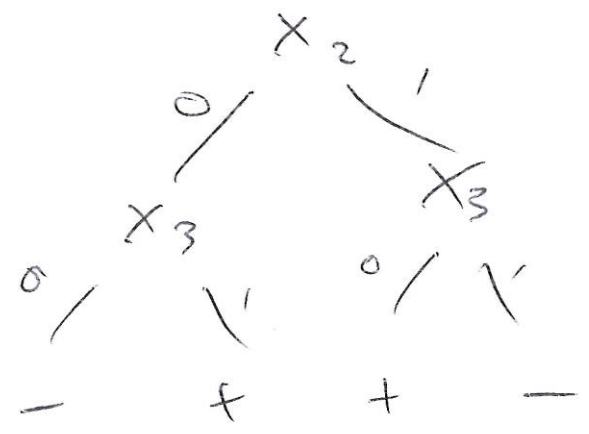
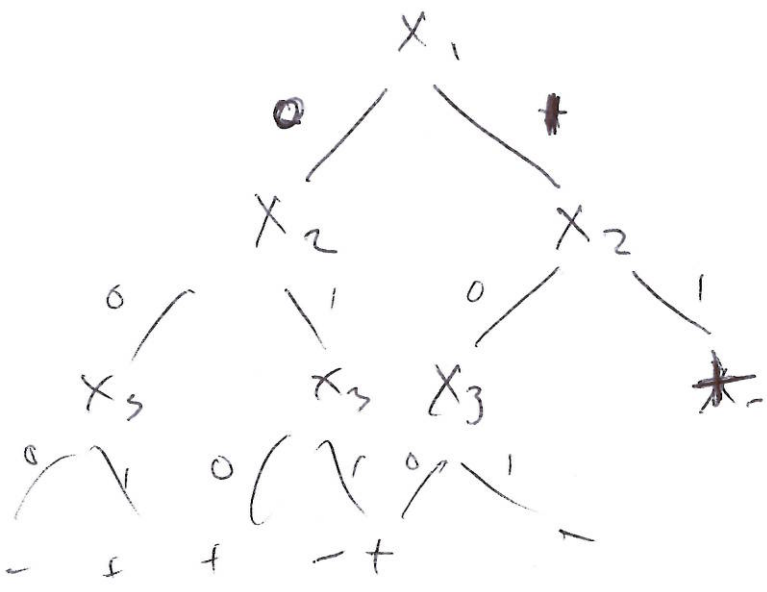
$$H(p) = -p \log p - (1-p) \log (1-p)$$



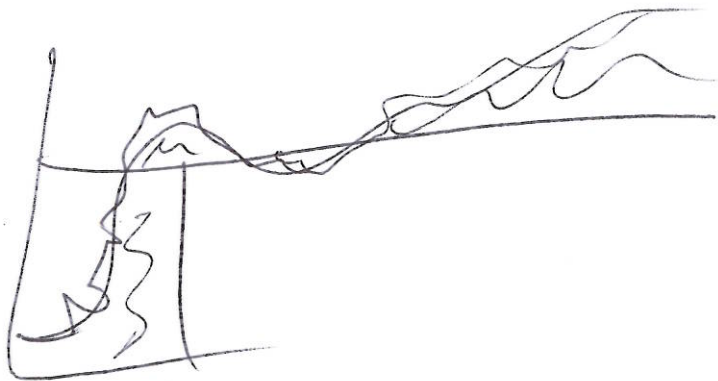
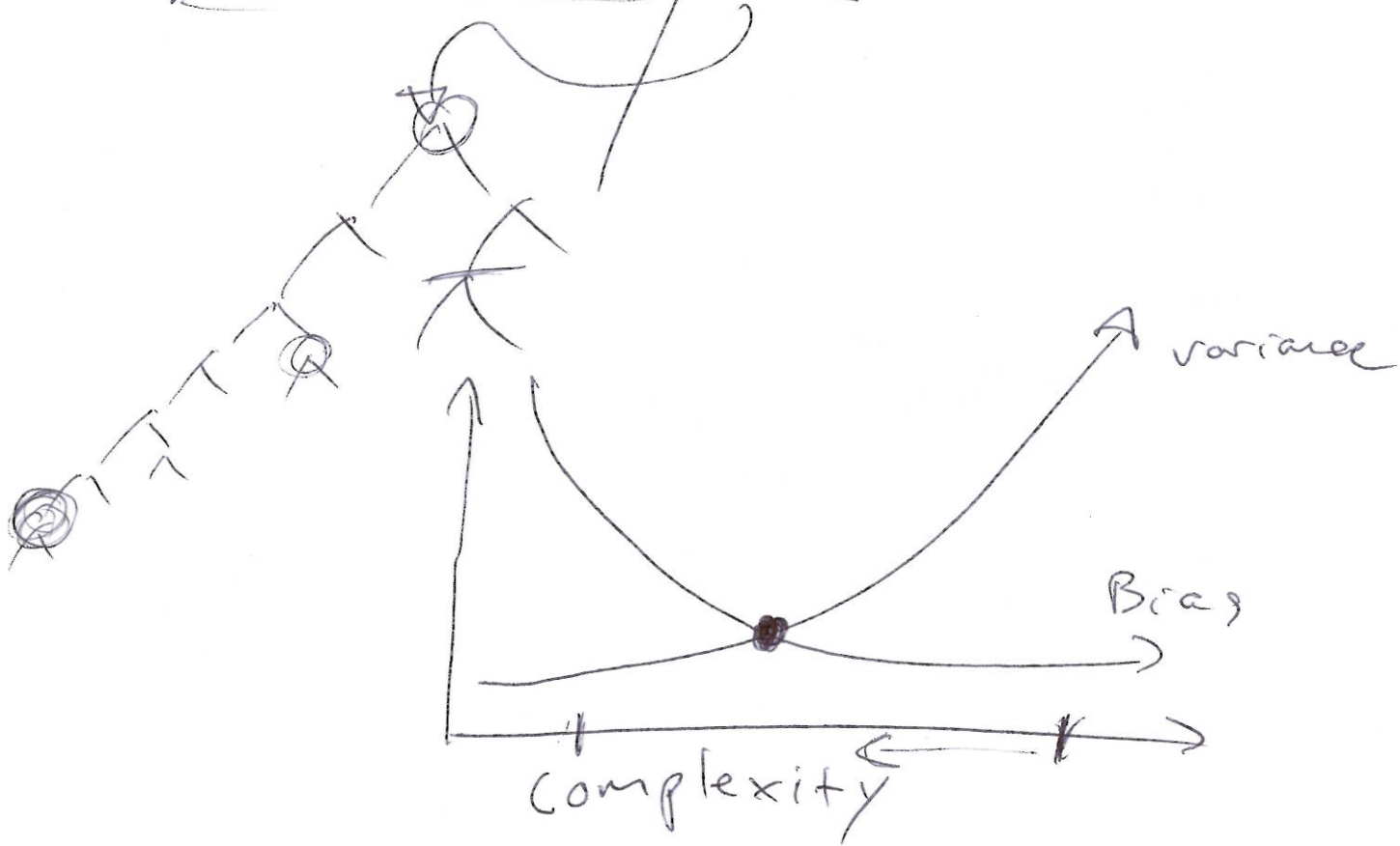
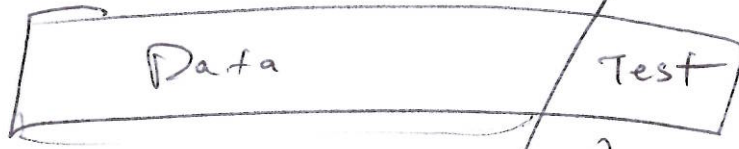
$$H(Y|X) = P(X_i) H(P(Y|X_i)) + P(\sim X_i) H(P(Y|\sim X_i))$$



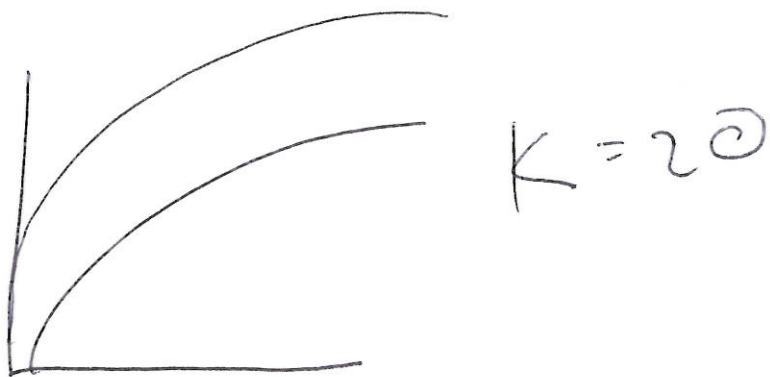
X_1	X_2	X_3	Y
0	1	0	+
1	0	1	+
1	0	1	+
1	1	0	+
1	0	0	-
0	1	1	-
0	0	0	-
0	1	1	-



they overfit.



K N N

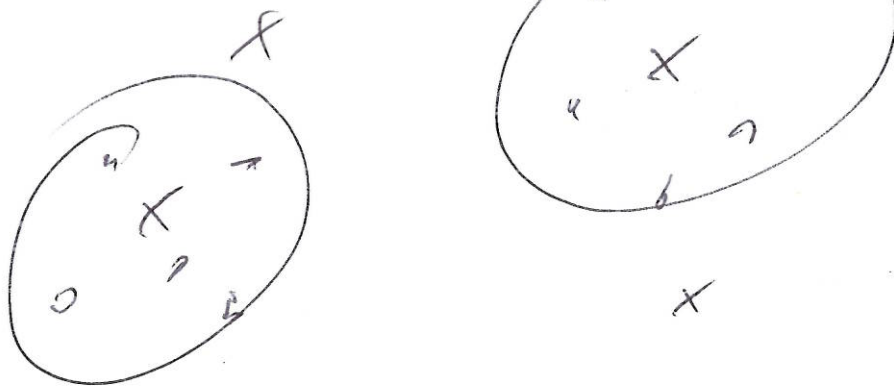


K-means

Find means

→ Reassign points

Repeat \times

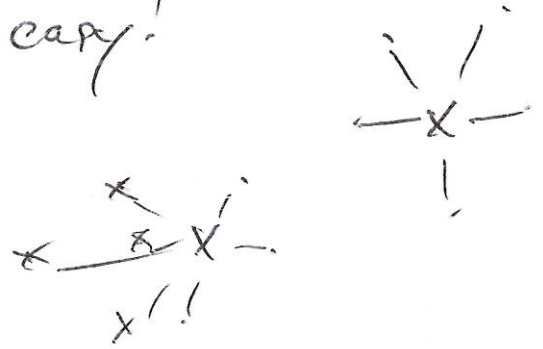


\times
4 major assumptions:

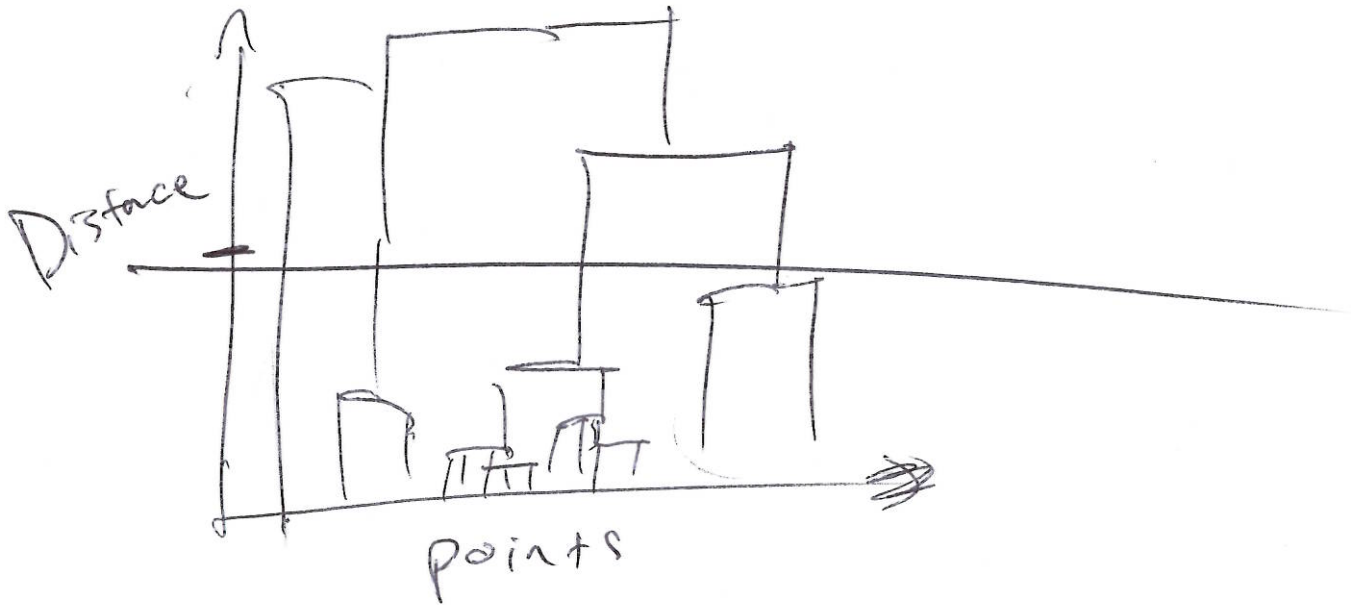
- there exist clusters
- Spherical clusters
- Clusters are equal size
- no outliers

minor assumption
- we know K

Choosing k is easy!

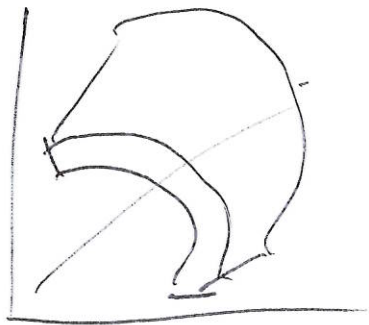
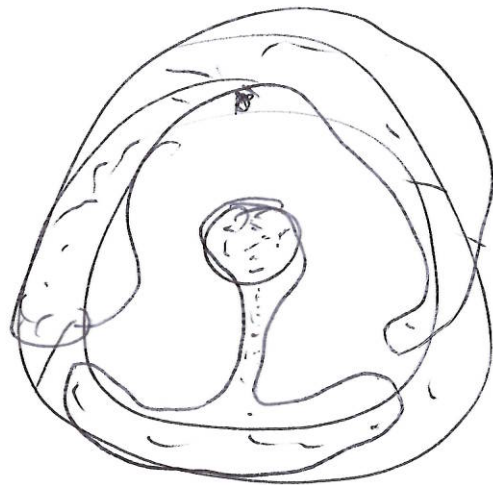
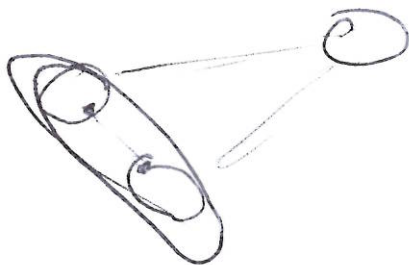


Dendrograms!



Hierarchical Clustering

Single link: merge clusters
based on closest two points



Complete link

farthest two points

