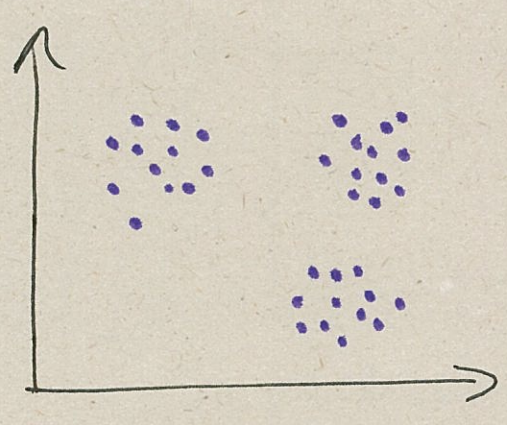
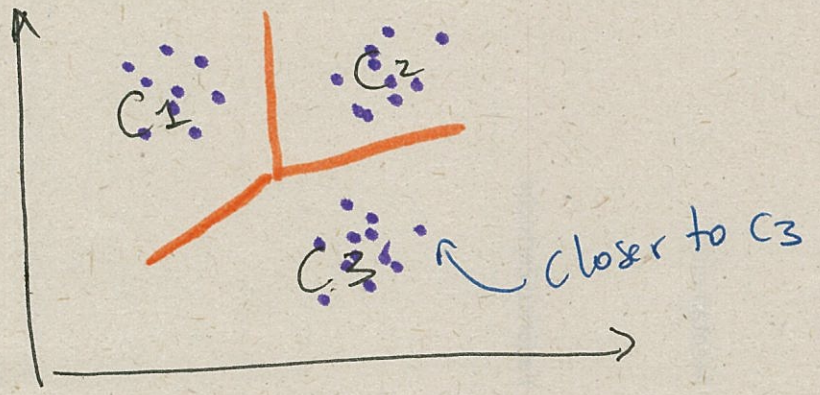
(k is given)



- Find $\rho = \{C_1, C_2, C_3\}$
- Form Voronoi tesselation



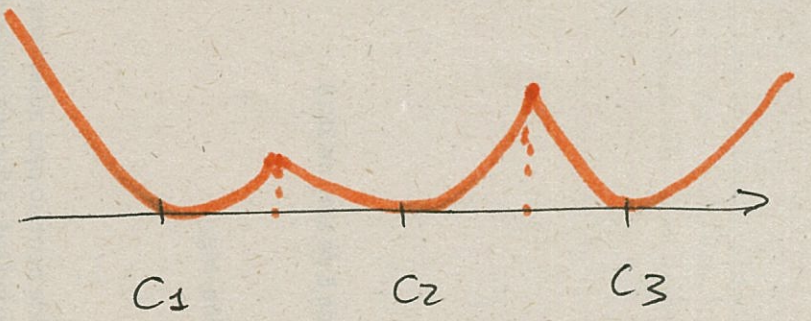
Stats Pt of View

$X_1, \ldots, X_n \sim P$

Given ρ:

$\text{loss} = \mathbb{E}\left[\min_j \|X - C_j\|^2\right]$

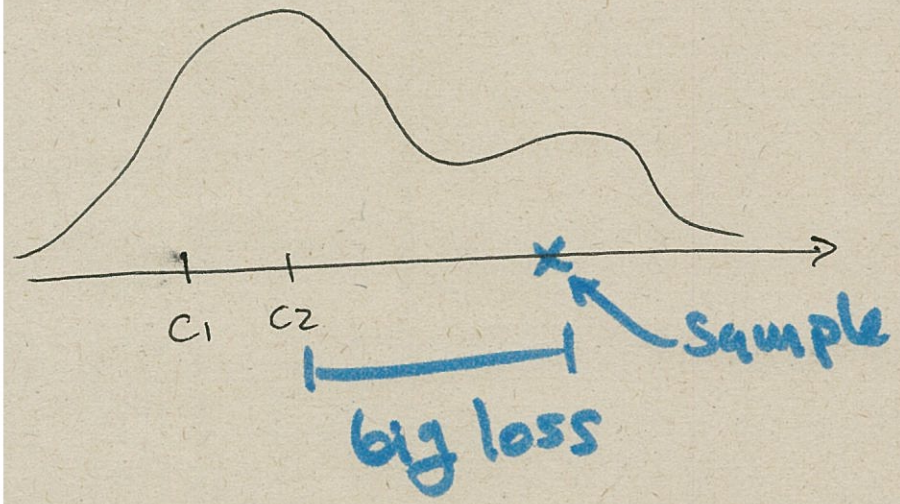
$R(C) :=$



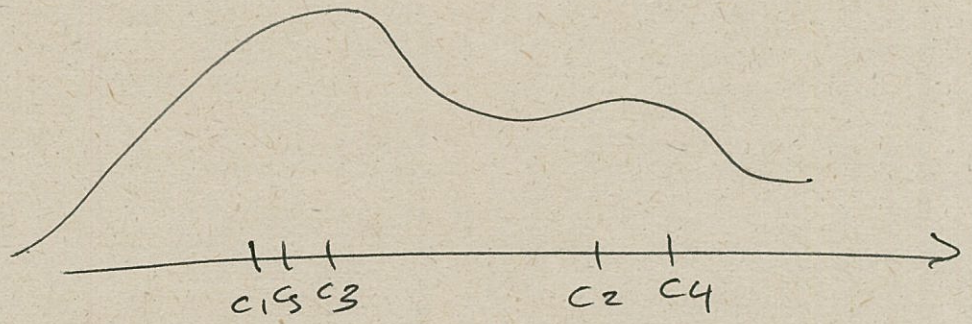
minimize $R(C)$
C

Non-convex optim problem
(In fact NP-Hard)

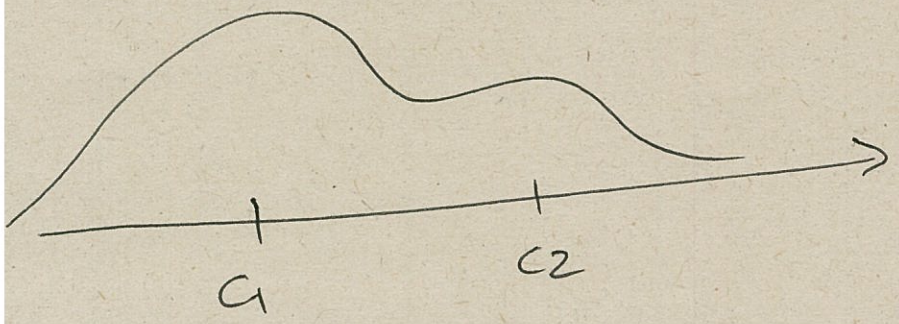
Choose c_1 & c_2:



5 c_js:



R(c) decreases with # of c_js.



Good Approx:

(Lloyd's algo)

① Choose $c_1, \ldots, c_k$.

② Form the clusters.

Compute $d_j(x_k) = \|x_k - c_j\|^2$

③ Reestimate centers:

$$c_j \leftarrow \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

Repeat until convergence.

(local min)