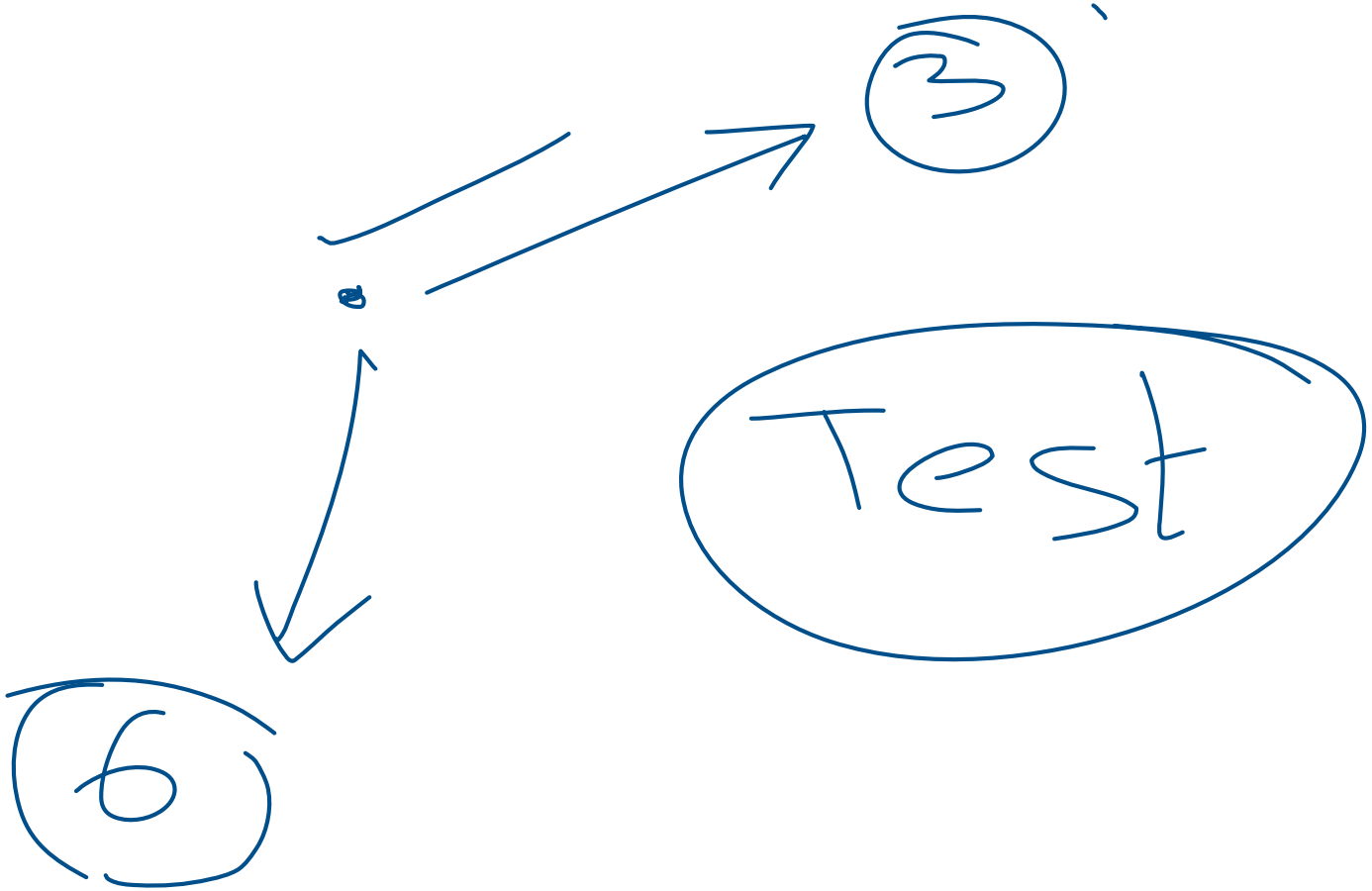
K - nearest Neighbours

- Supervised algorithm

Age	Gr	Label
—	—	1
—	—	0

Train Test

— Euclidean Distance

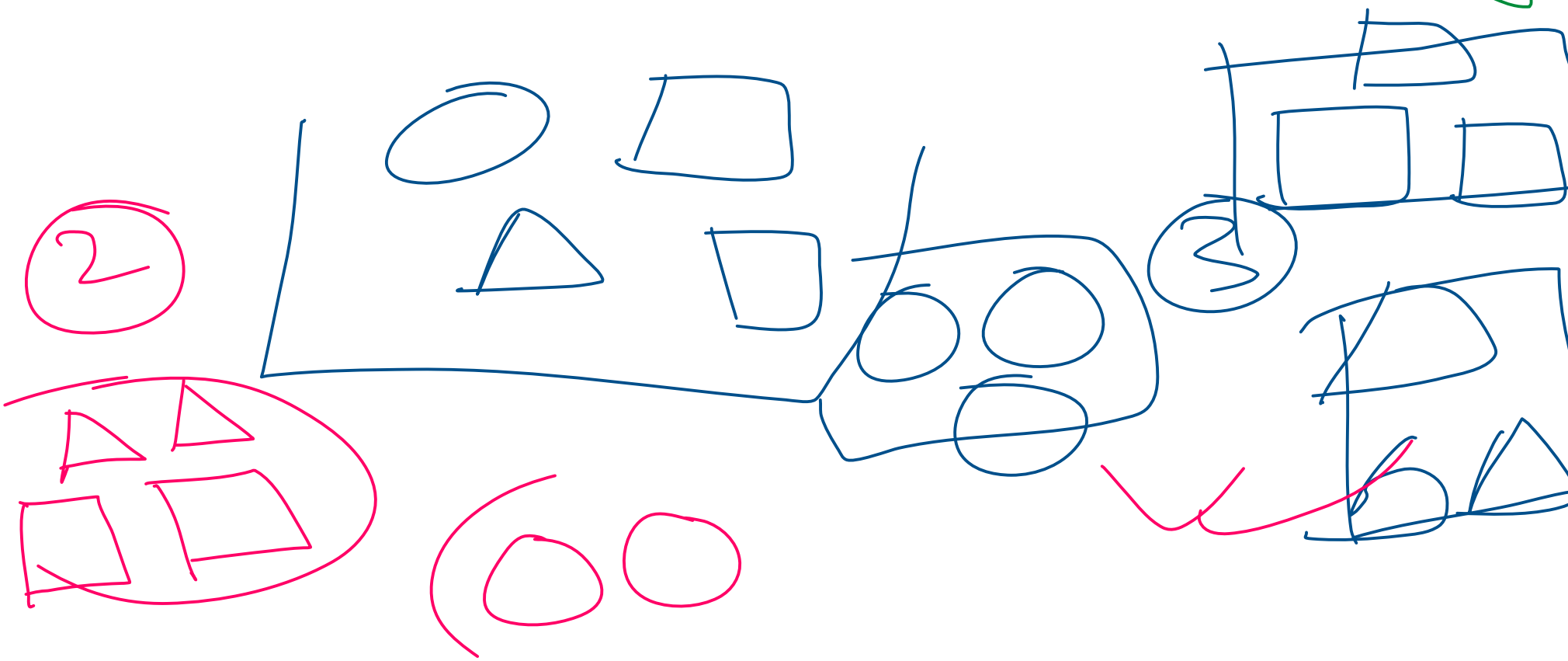


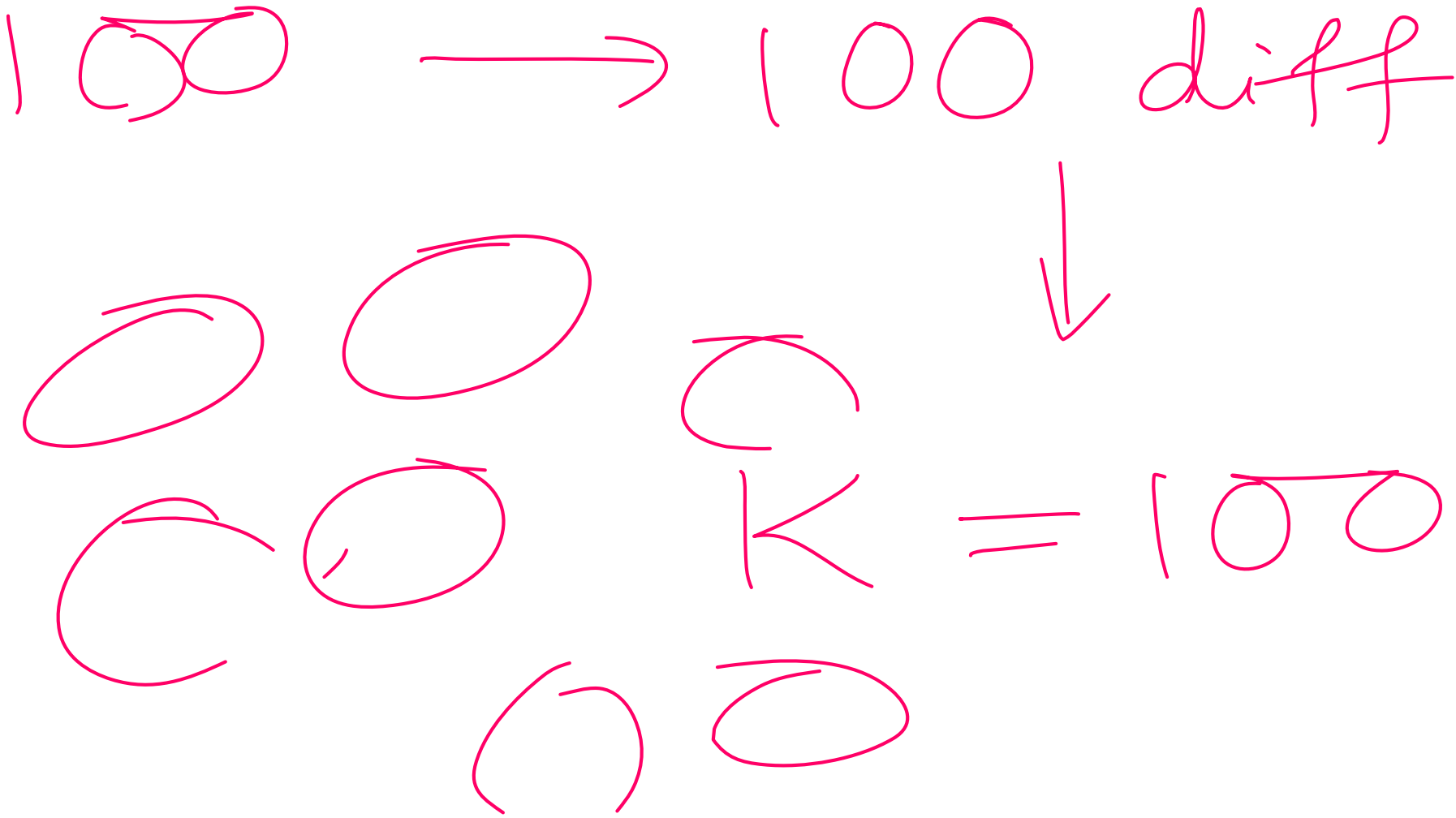
$$E.D. = \sqrt{\sum_{i}^{n} (x_i - y_i)^2}$$

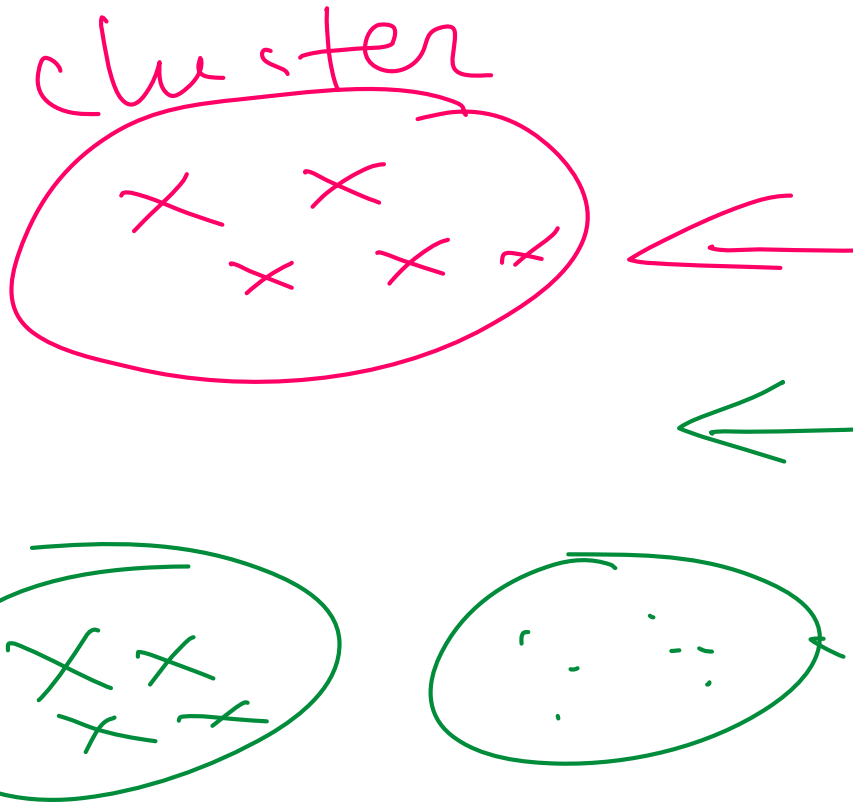
3 steps!

1. Calc. Eucl. Dist.
2. Find nearest neigh
3. Make Predictions

K-Means → clustering




100 → 100 diff
K = 100


cluster
1st Property
2nd ,,

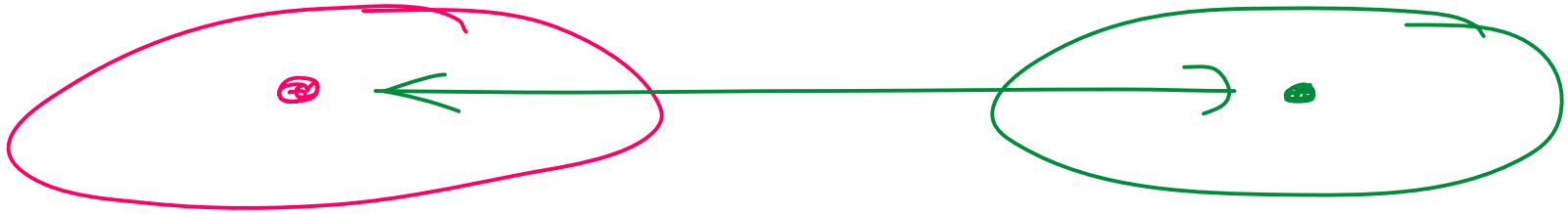
- Clustomer
- Document
- Image Segmentation
- Recommendation

Inertia

- evaluates first property

Sum of distances of all points from Cent of the clust

Dunn Index



$$DI = \frac{\min(\text{inter cluster dis}}{\max(\quad,,)}$$

Silhouette score & plot

- K means → soln
- evaluation

K - Means | Steps

1. Choose #K [no of cluster
2. k random points

③ Assign points to respective clusters

Avg

Avg

④ C_1, C_2 changed. Recompute.

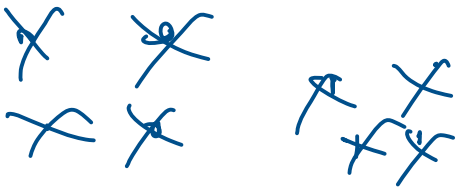
⑤ Step 3, 4. → Repeat

until C_1, C_2 → constant

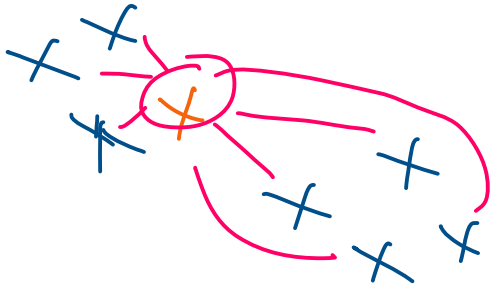
k-means++

1. First cluster is chose randomly. Just one.
2. $D(x)$ → all data point
3. New cluster centre → $P(x) \propto D^2(x)$

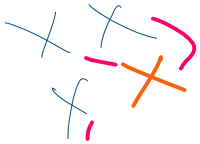
④ Repeat 2, 3 until k clusters.

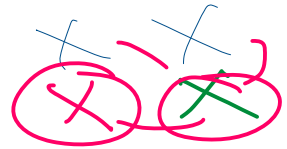


S-1



S-2

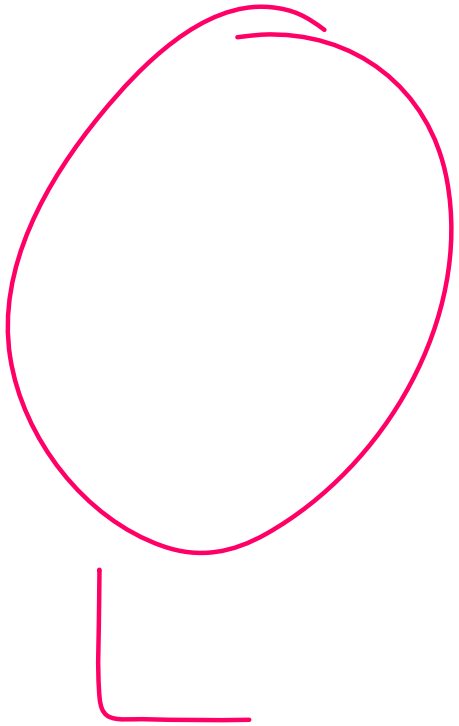


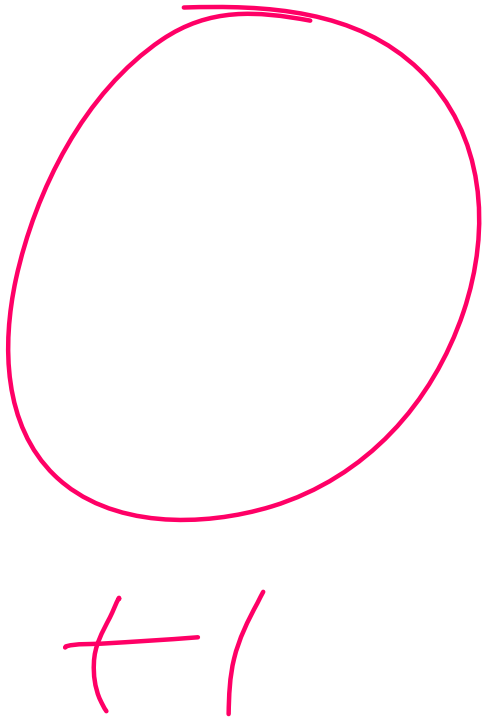


S-3

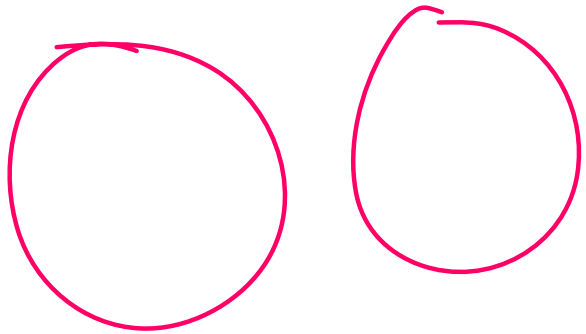
(Repeat)

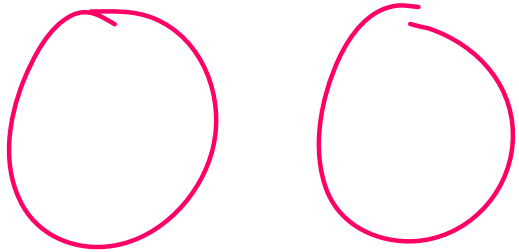
- Income & Debt

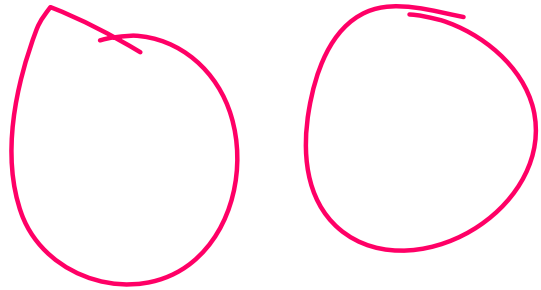


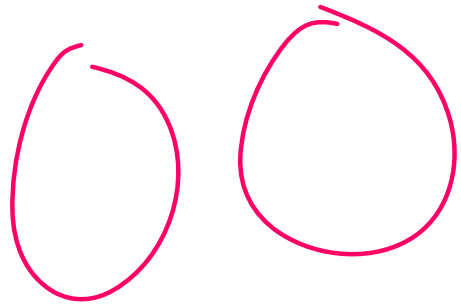


+1









Elbow Curve