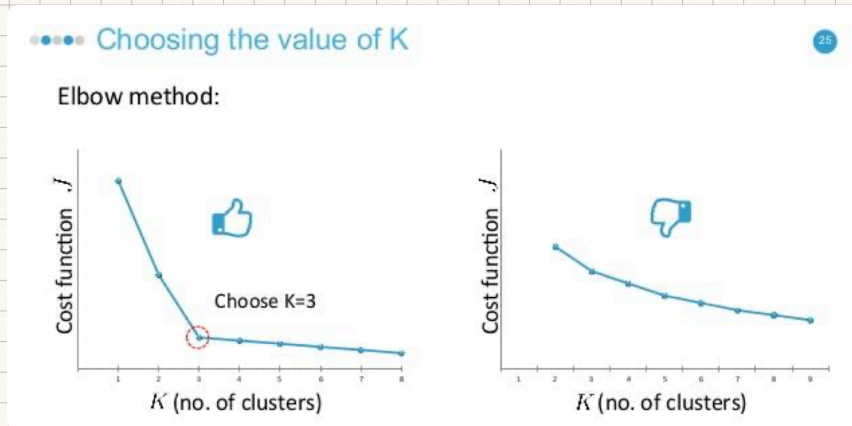


ELBOW METHOD

Aumentare il numero di cluster può migliorare la varianza dei punti dentro un singolo cluster. Questo però non succede ogni volta che si aumenta il numero di cluster, anzi la situazione può arrivare a saturarsi. Per questo bisogna trovare un K ottimale.

Basicamente questo metodo consiste nel testare per ogni K la **somma delle varianze** di ogni cluster e esaminare il grafico al variare di K :



Scegliere poi il valore di K che presenta il più significativo punto di svolta (**Turning Point**) per cui dopo ciò le varianze non migliorano troppo.

CROSS VALIDATION METHOD

Semplicemente si divide il dataset in M parti e poi utilizzare $(M-1)$ samples per creare i cluster e usare l' M -esimo sample per testare i clusters trovati.

Come testare un cluster:

Per ogni punto del test-set valutare la distanza dal centroide più vicino e poi usare la somma del quadrato delle distanze fra tutti i punti del test-set per misurare quanto bene il modello fitted il test-set.

Ripetere la procedura M volte per ogni K che si vuole testare.