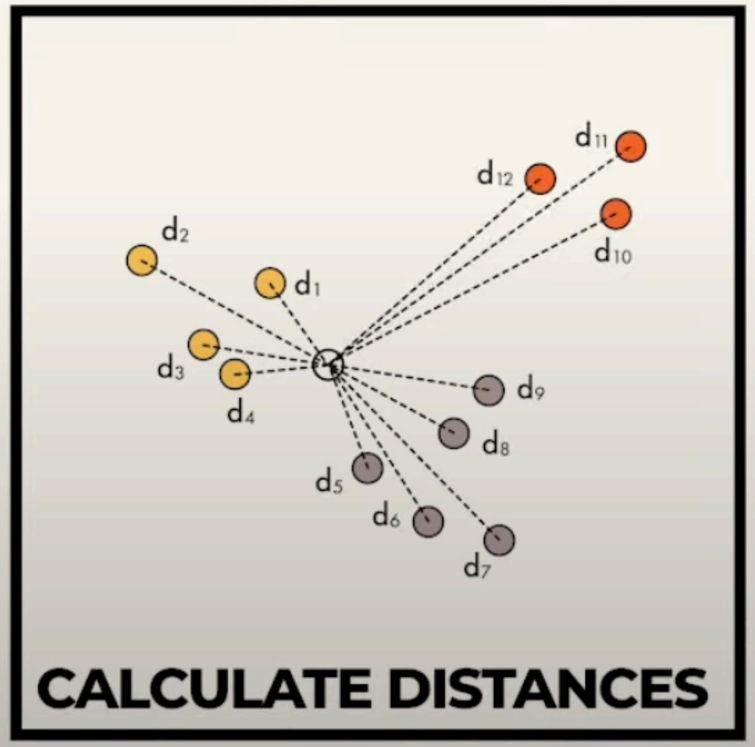
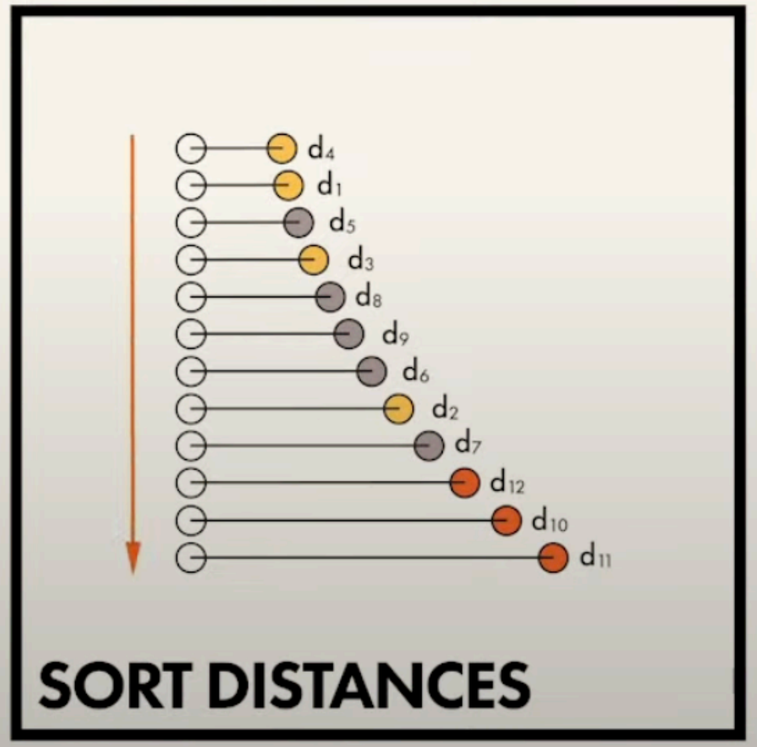
მანგის გაყიდვების ანალიზი

KNN & SVR

თორნიკე კიკაჩეიშვილი

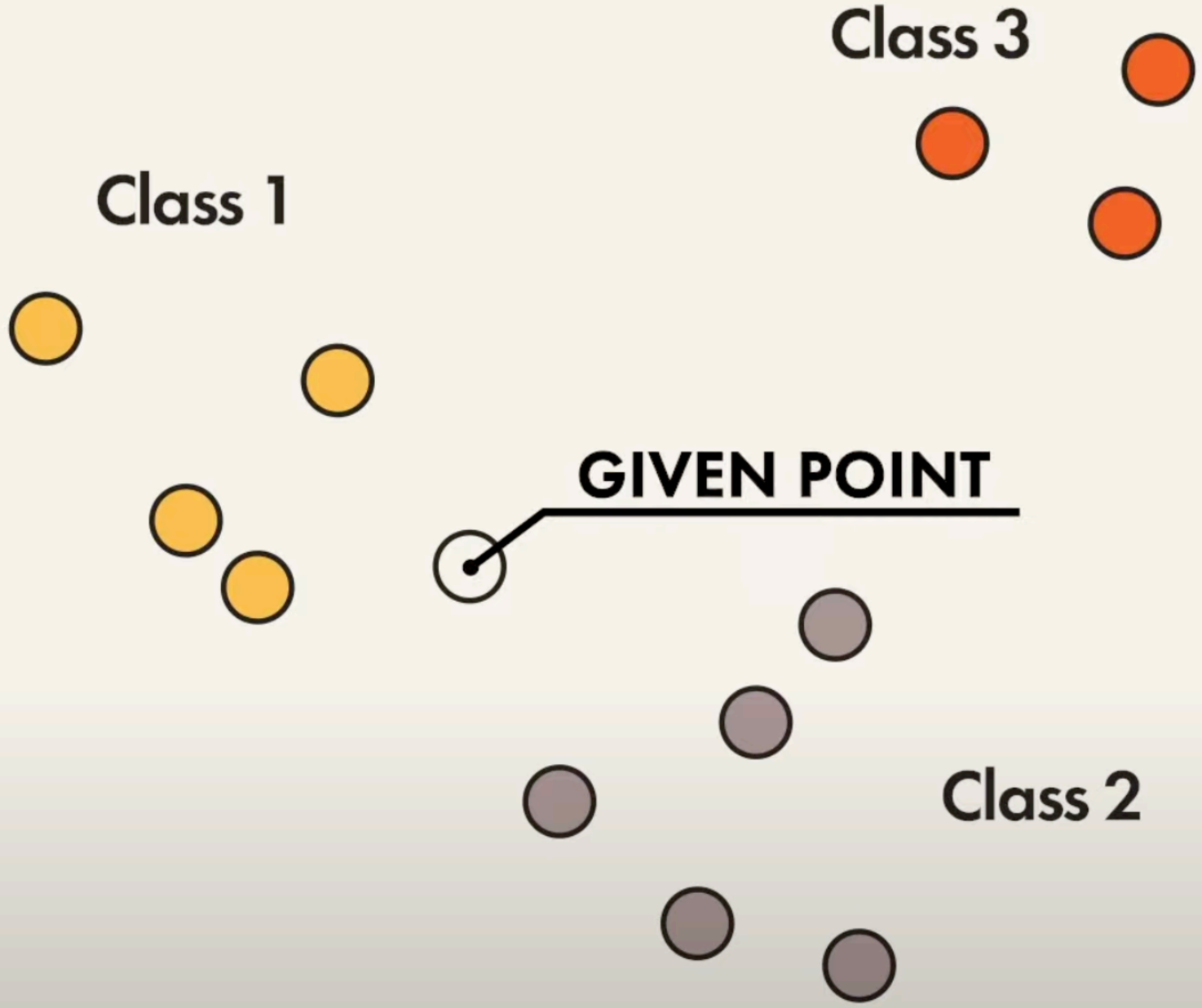
K NEAREST NEIGHBORS





STEP 1

Initialization



STEP 2

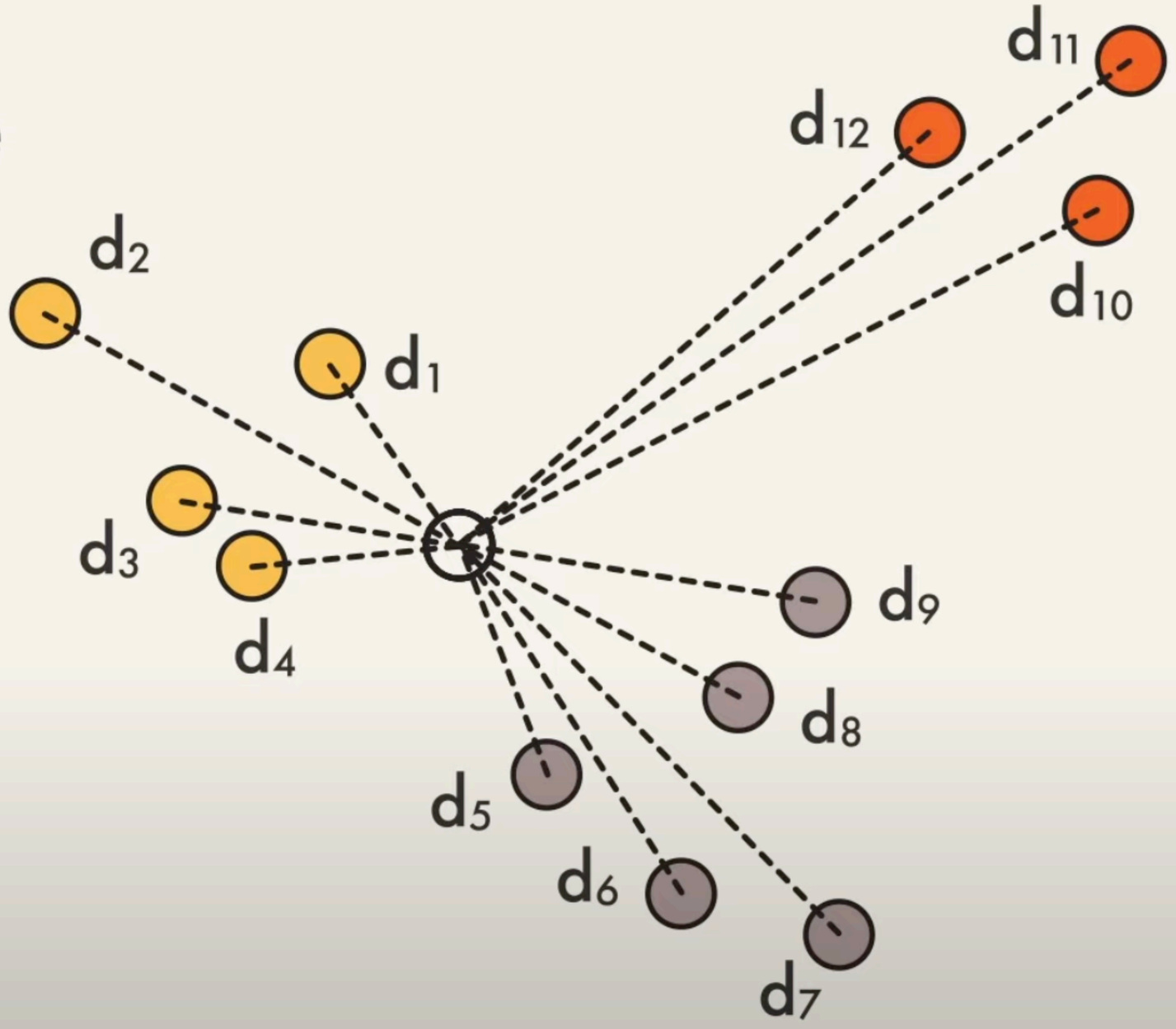
Calculate Distance

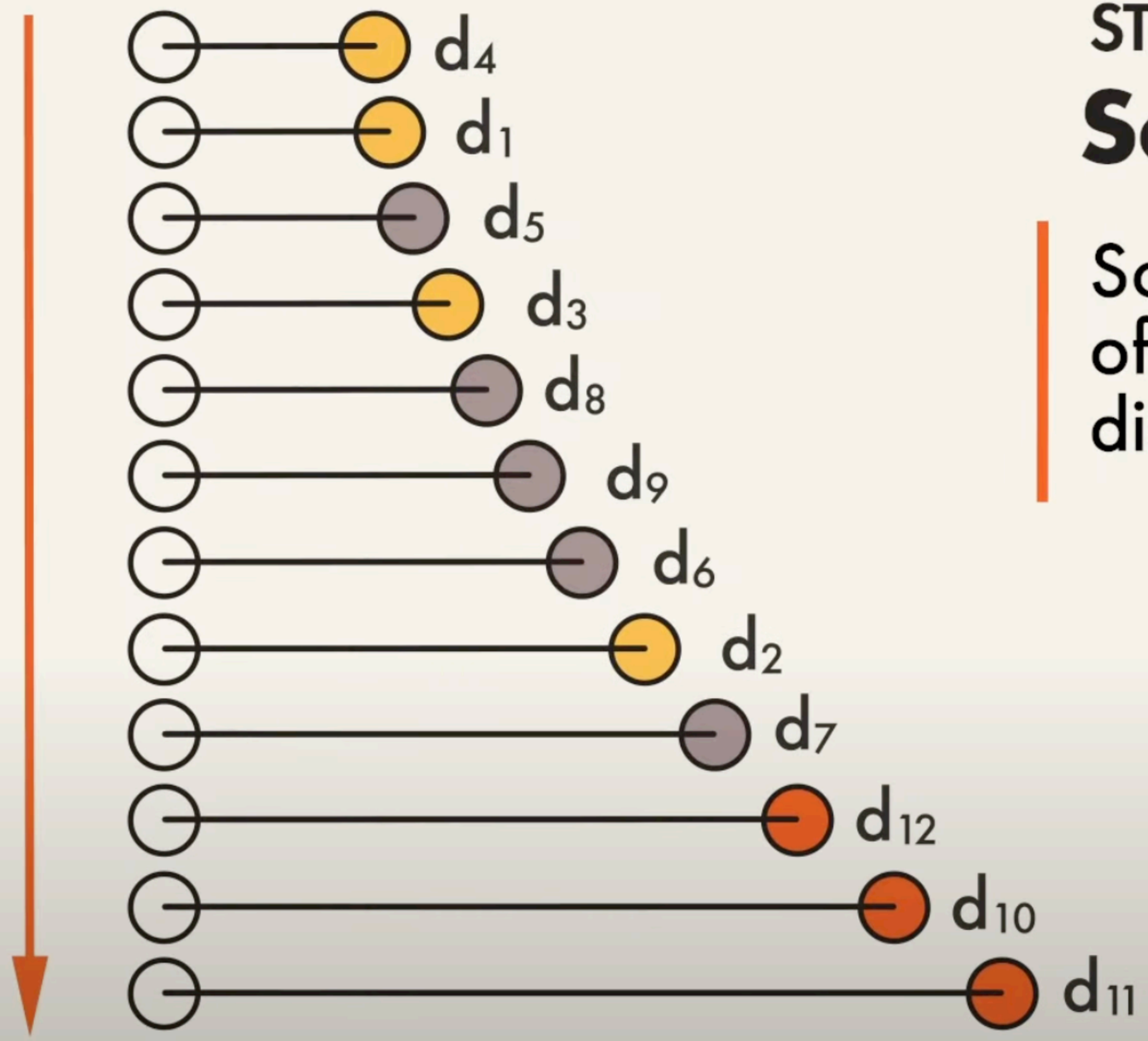
Euclidean:

$$d(x,y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2}$$

Manhattan/city - block:

$$d(x,y) = \sum_{i=1}^{m} |x_i - y_i|$$

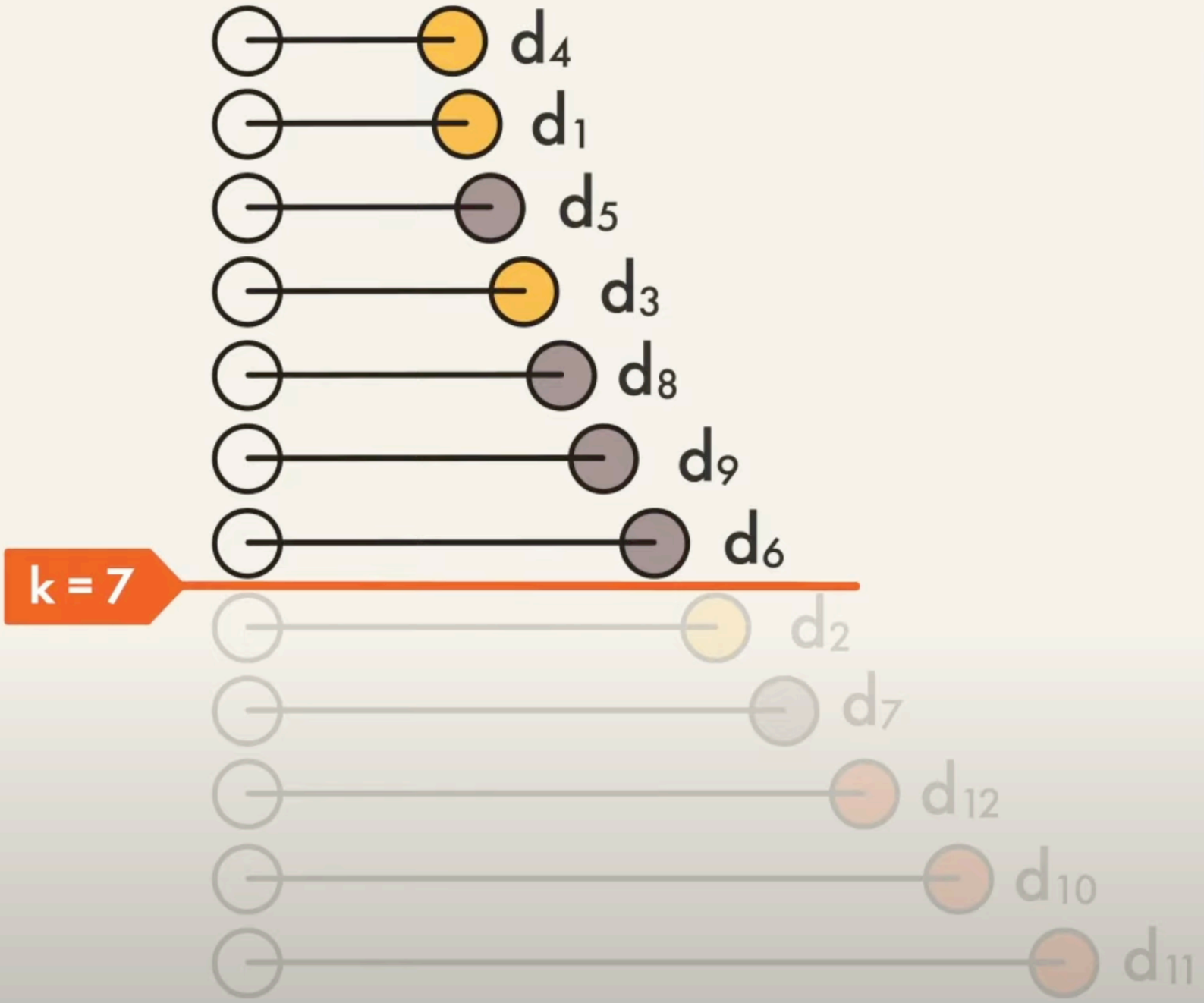




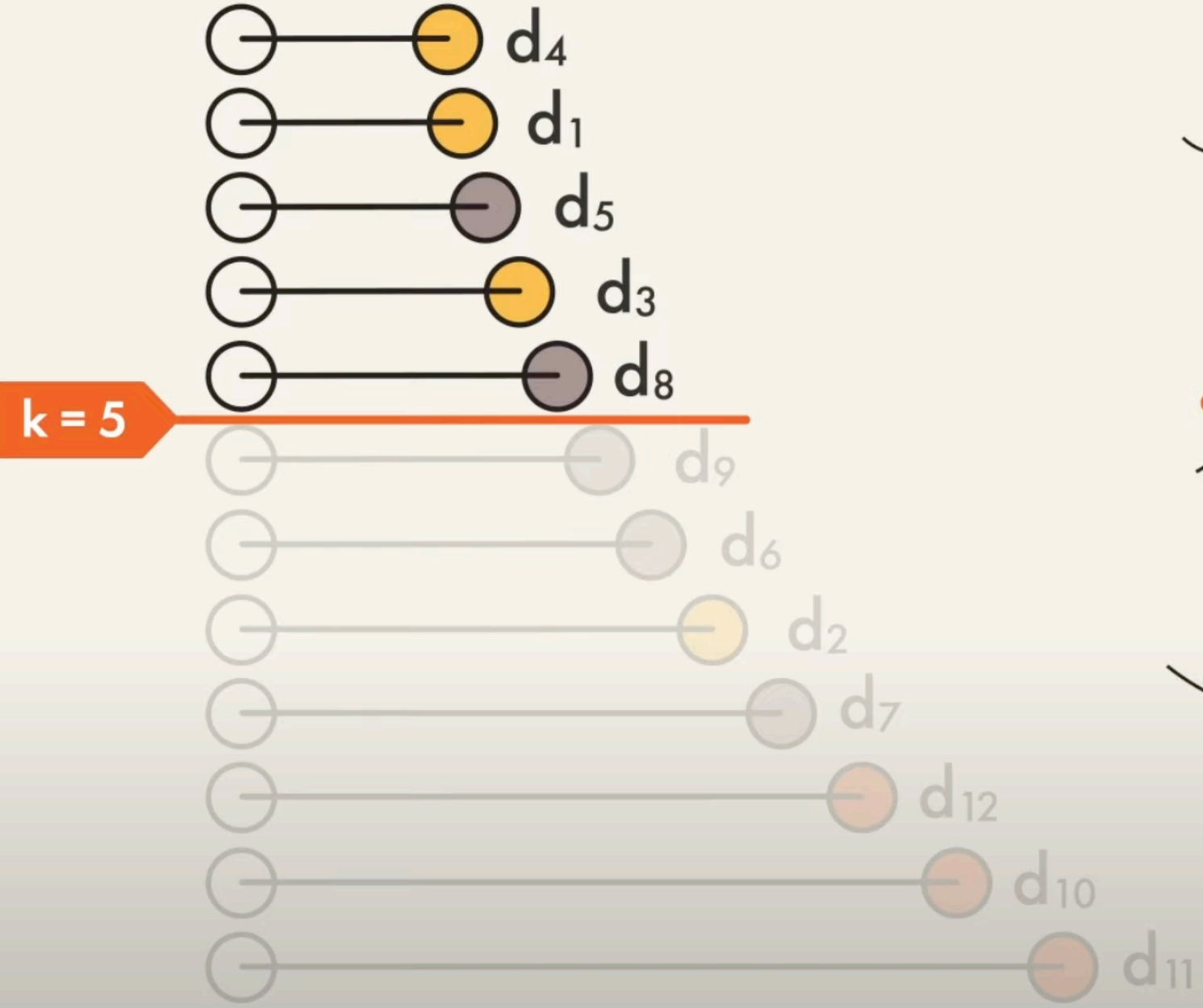
STEP 3

Sort Distance

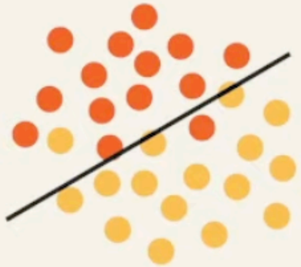
Sort the nearest neighbors of the given point by the distances in increasing order



Classification



Small K
Low Bias; High Variance
Overfitting



Large K
High Bias; Low Variance
Underfitting

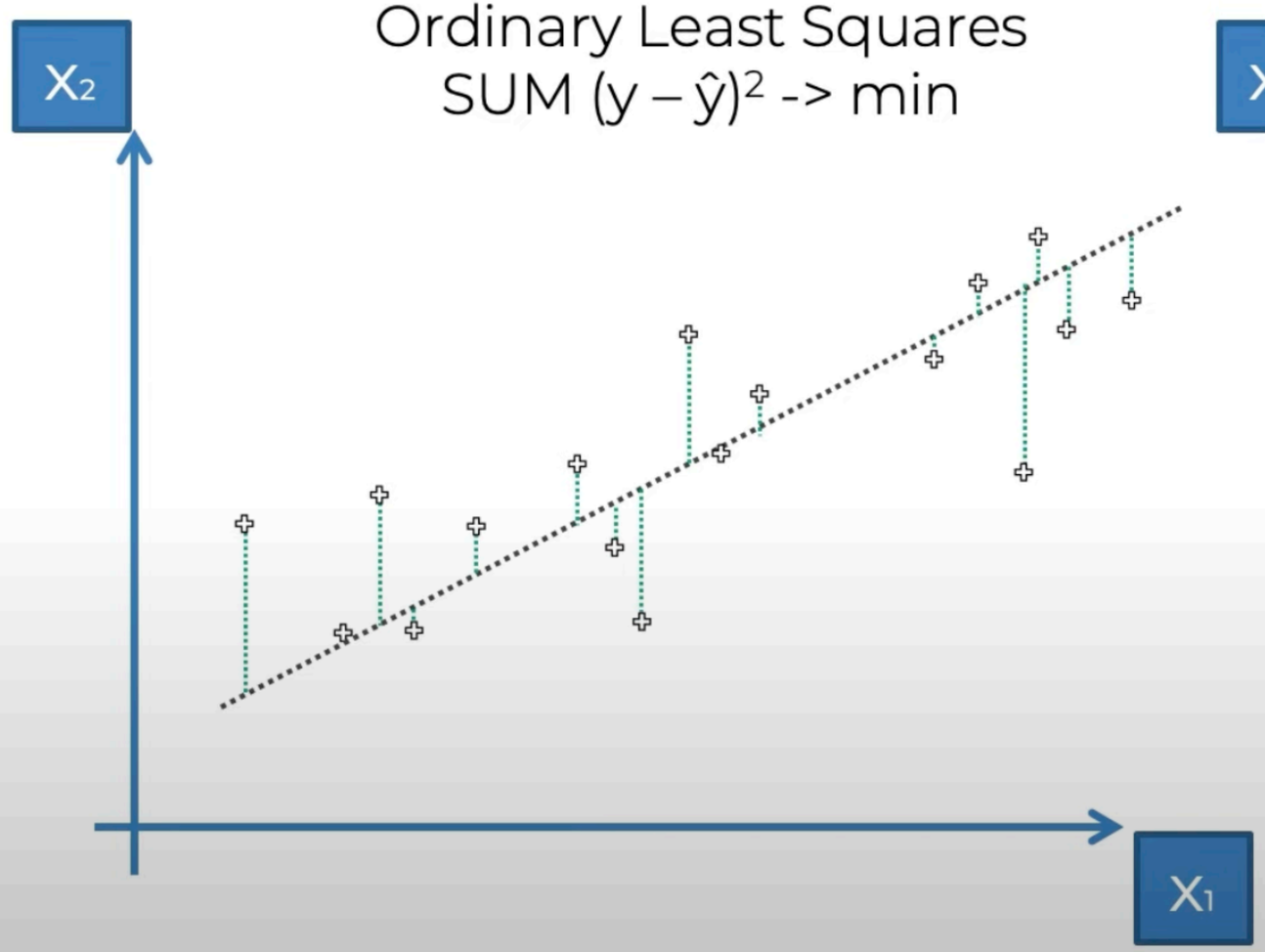
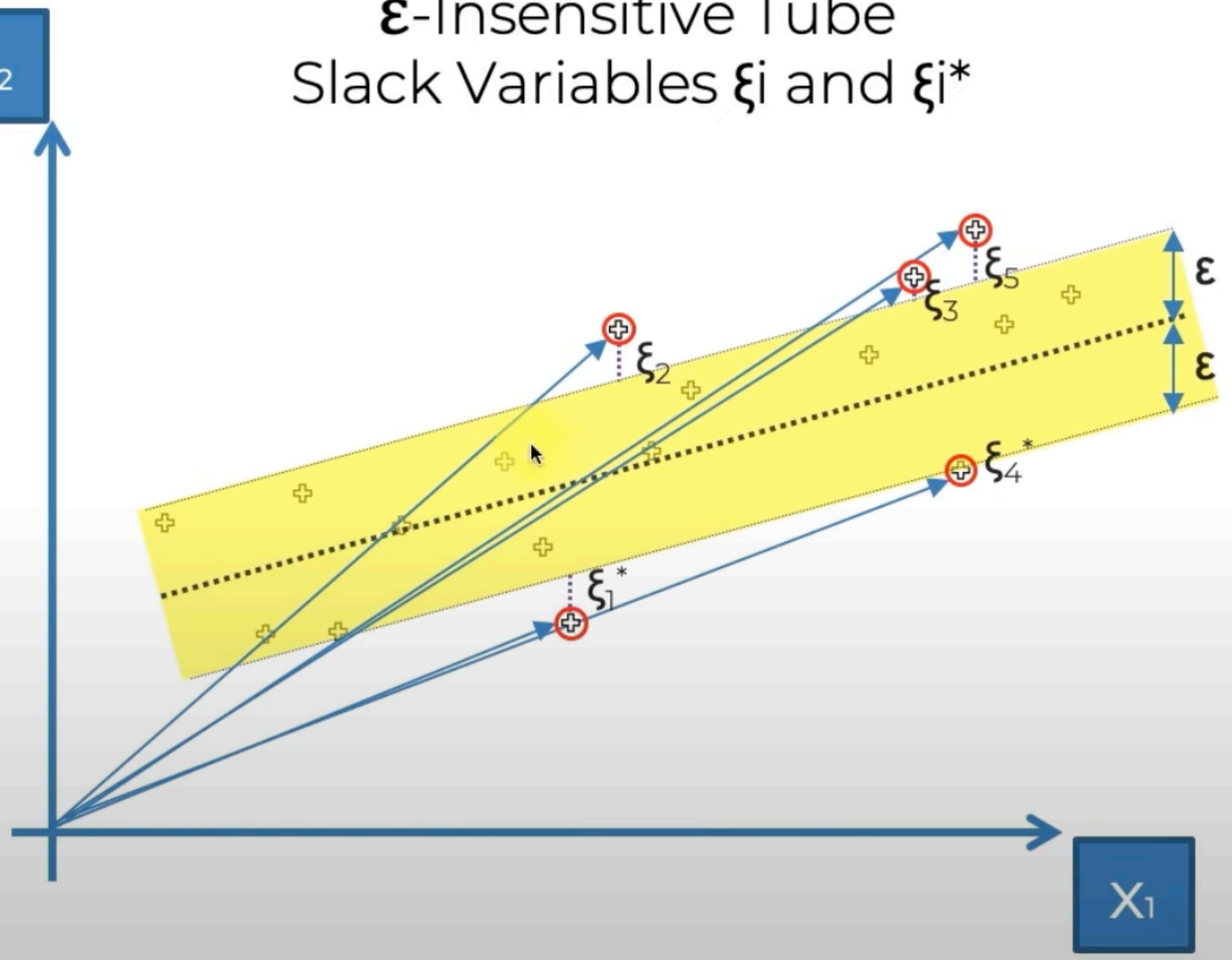
Best K value controls
the balance between
overfitting and underfittin

I like this

SVR Intuition

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \rightarrow min$$

Ordinary Least Squares
SUM $(y - \hat{y})^2$ -> min



ε-Insensitive Tube
Slack Variables ξi and ξi*