



K- means discretisation

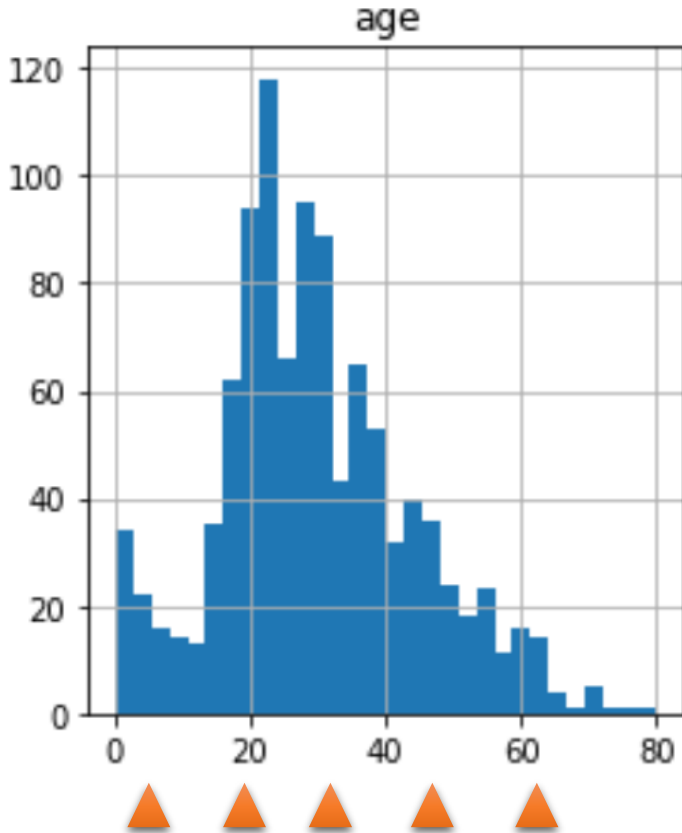


K-means discretisation: definition

This discretisation method consists in applying k-means clustering to the continuous variable.

More details about k-means [here](#).

K-means discretisation: calculation



Intervals = 5

Find 5 centroids

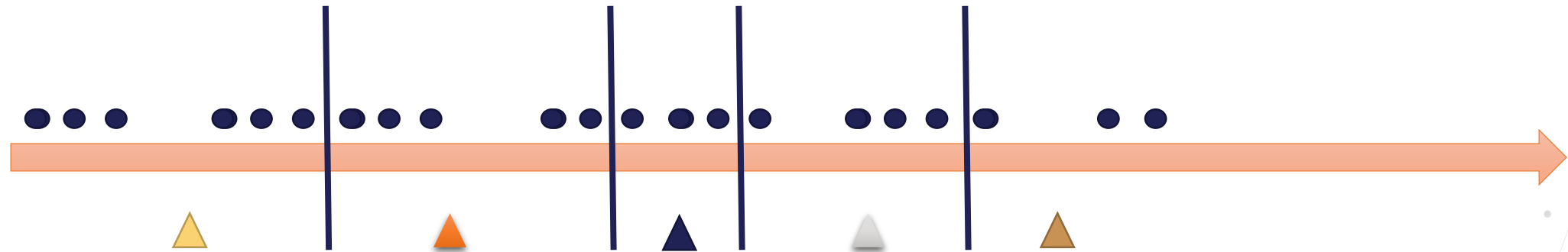
Divide data into clusters according to distance to centroids

K-means discretisation: calculation

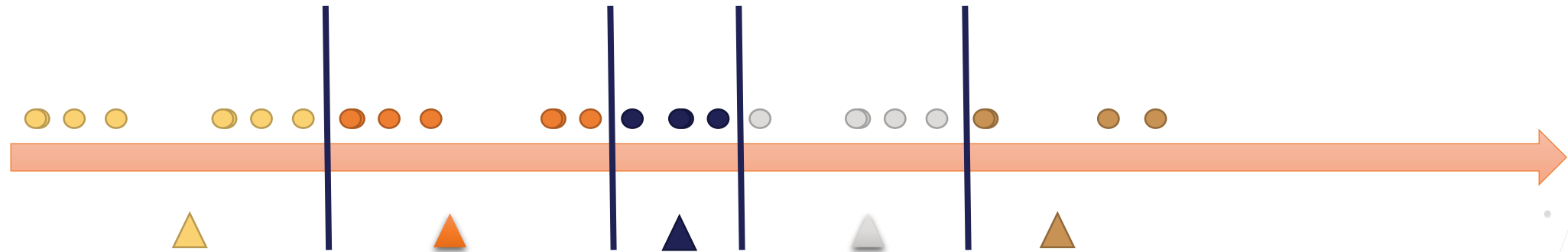


Age

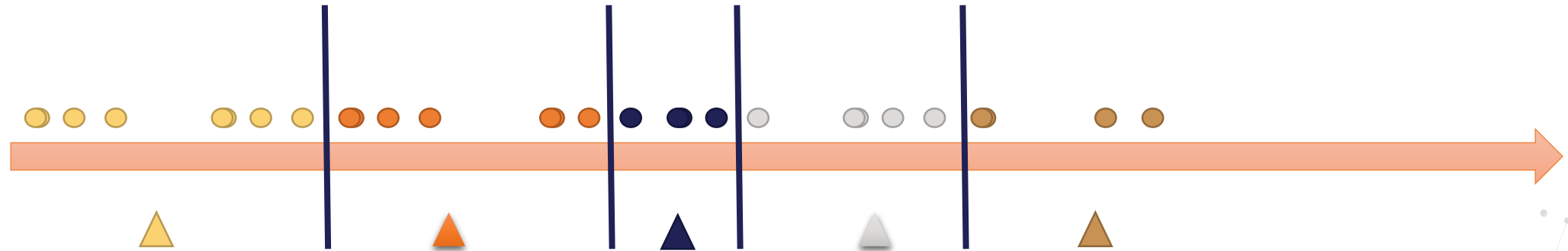
K-means discretisation: calculation



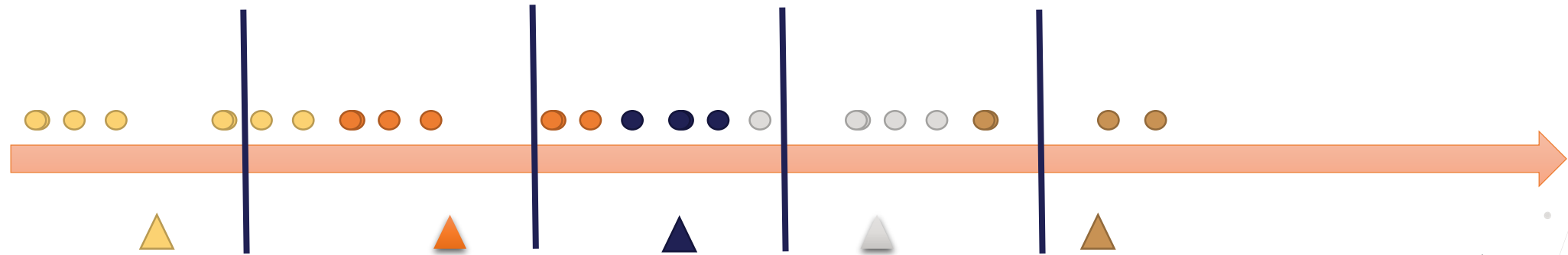
K-means discretisation: calculation



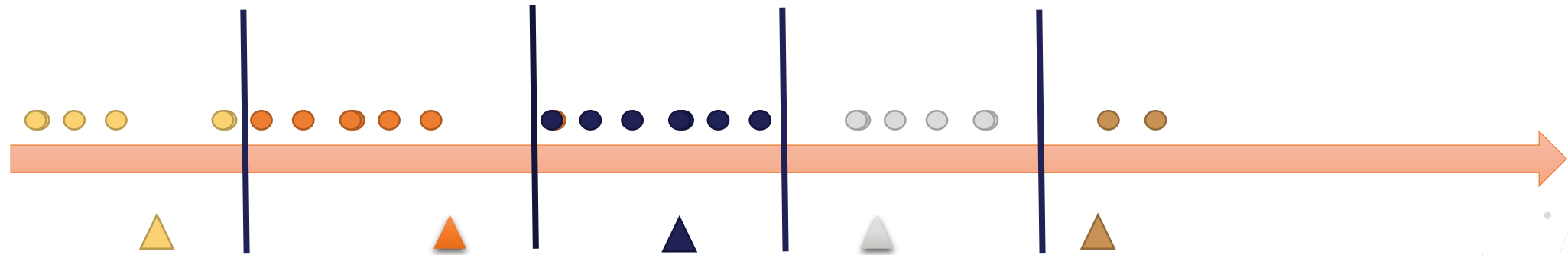
K-means discretisation: calculation



K-means discretisation: calculation

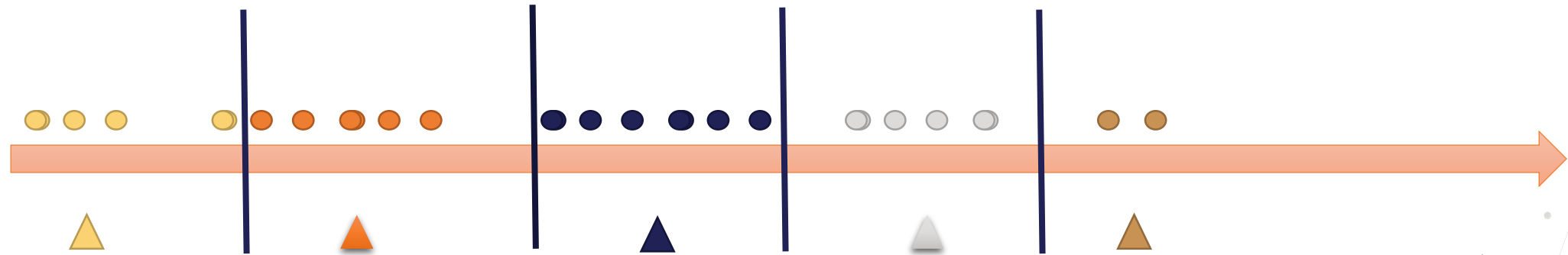


K-means discretisation: calculation



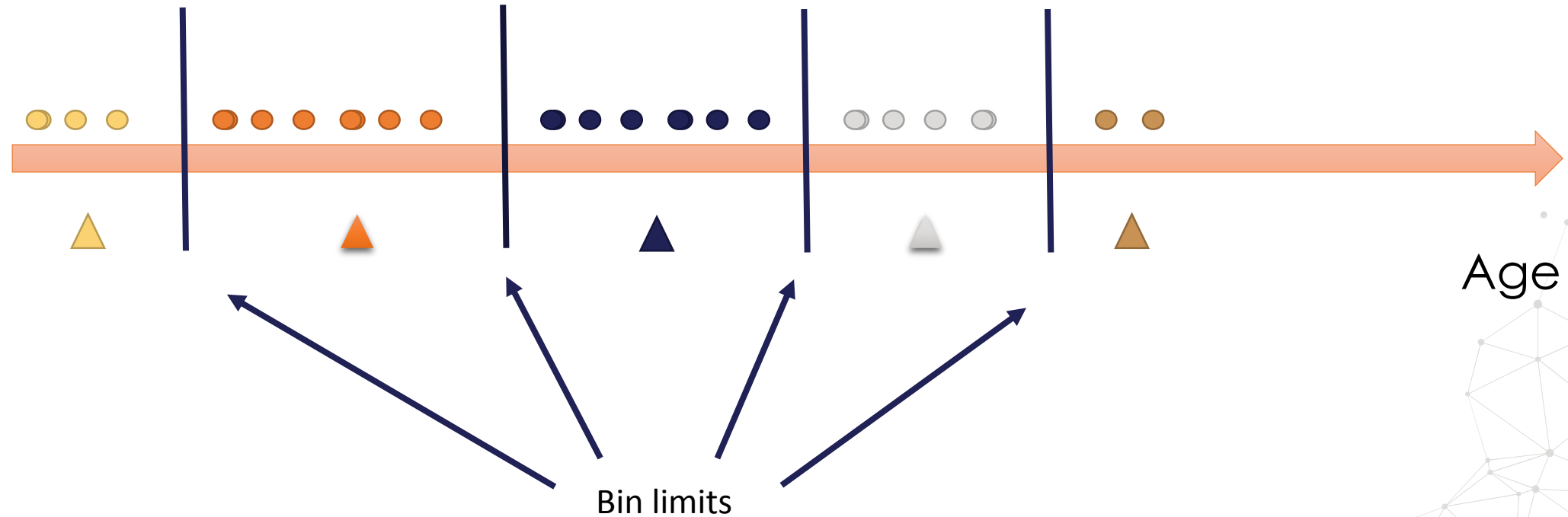
Age

K-means discretisation: calculation

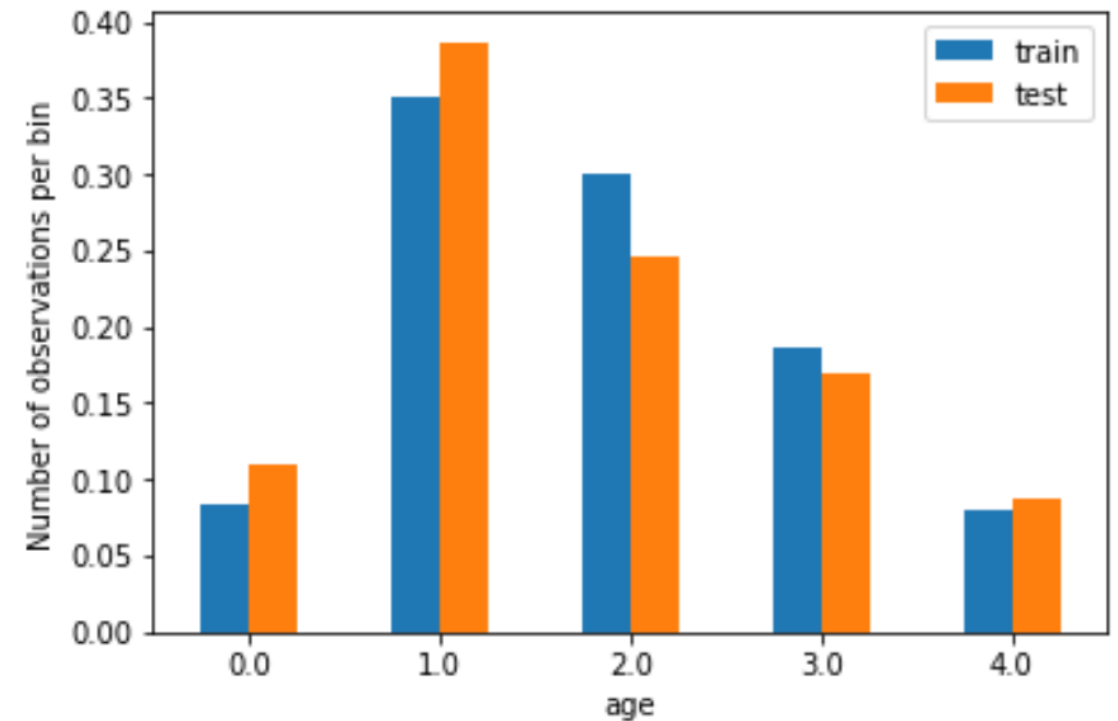
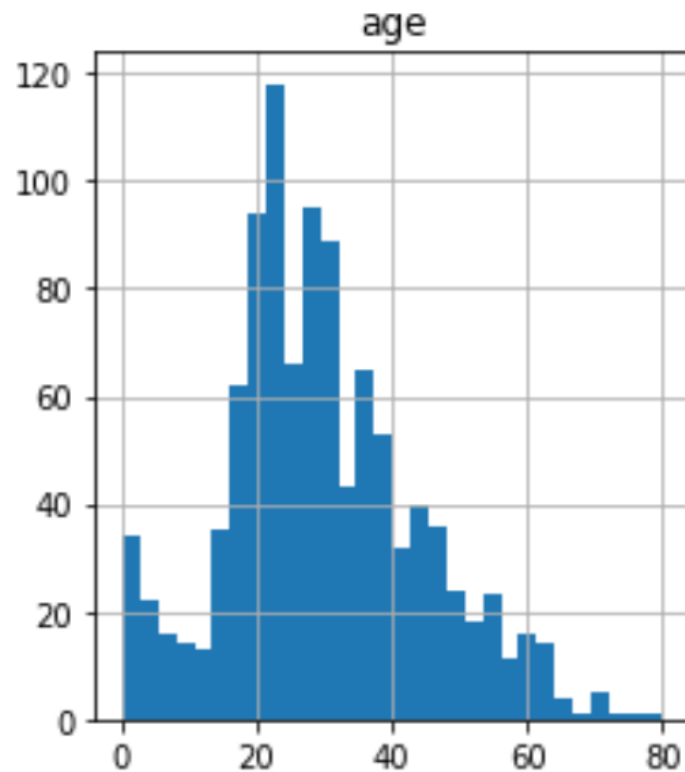


Age

K-means discretisation: calculation



K- means discretisation: definition



K-means discretisation: summary

- Does not improve value spread
- Handles outliers, although outliers may have an influence in the centroid
- Creates discrete variable
- Good to combine with categorical encodings

THANK YOU

www.trainindata.com