Evan Pradipta Hardinatha-1103204160

Principal Component Analysis

Pca is a method for compressing a lot of data into something that captures the essence of the original data.

Dimensions:

1 cell = 1D graph 2 cell= 2-D 3 cell=3D

Flatten : flatting 2D-1D 3D-2D

PCA takes dataset with lot of dimensions and flatten it to 2-3 dimension

PC1 Captures the most variation

PC2 Captures 2nd most

KNN

Step 1 : dataset with known categories, cluster with PCA

Step 2: add new cell, with unknown category to the PCA plot

Step 3 : Classify the new cell

If the new cell between two or more categories, simply pick categories with most votes

If votes are equal we can add to one of the categories or simply not inputting it to categories.

K must be try out before settling on one

Low K can be noisy

High K smooth over things, but you don't want to be so large with few samples category

Decision Tree

Decision tree makes a statement true or false

Classification: start on the top and go down until you get to the leaf

Top of the tree : root, then branches that have arrow to bottom, last leaf

Gini Impurity: 1(prob of yes)^2-(prob of no)^2