

Practice Problem 3

Ian Chen

February 14, 2024

Problem 1 — Using a KNN classifier with $k=3$, predict whether the following tissue paper will be good or bad for your science experiment. *Do not worry about normalizing the data.*

Answer

$$d(p, q): \sqrt{\sum_i (p_i - q_i)^2}$$

Attributes: Color-White, Color-Yellow, Color-Green, Acid Durability, Strength

$$d(5, 1): \sqrt{(1-0)^2 + (0-1)^2 + (0-0)^2 + (4-7)^2 + (6-7)^2} = 3.464101615$$

$$d(5, 2): \sqrt{(0-0)^2 + (0-0)^2 + (0-0)^2 + (4-7)^2 + (6-4)^2} = 3.605551275$$

$$d(5, 3): \sqrt{(1-0)^2 + (0-1)^2 + (0-0)^2 + (4-3)^2 + (6-4)^2} = 2.645751311$$

$$d(5, 4): \sqrt{(1-0)^2 + (0-0)^2 + (0-1)^2 + (4-1)^2 + (6-4)^2} = 3.872983346$$

The three closest neighbors are 3, 1, and 2.

The majority of the three closest neighbors are bad.

∴, the tissue paper is bad for the science experiment.

Problem 2 — Now use weighted voting to predict if the tissue paper is good or bad. Does the answer change?

Answer

$$\text{weight factor: } w = \frac{1}{d^2}$$

$$w(1) = \frac{1}{3.464101615^2} = 0.08333333334$$

$$w(2) = \frac{1}{3.605551275^2} = 0.07692307694$$

$$w(3) = \frac{1}{2.645751311^2} = 0.1428571429$$

$$w(4) = \frac{1}{3.872983346^2} = 0.06666666667$$

Three closest neighbors are 1, 2, and 4.

$$\text{weight of bad} = w(1) + w(2) = 0.1602564103$$

$$\text{weight of good} = w(4) = 0.06666666667$$

The weighted majority of the three closest neighbors are bad.

∴, the tissue paper is bad for the science experiment, and the answer is the same as the unweighted one.