KNN – Final Remarks

Dr. Muhammad Wasim

Feature Scaling

- Different scaling can affect the classification results. So, scaling the numerical features is necessary.

Example

Person	Height(inches)	Height(centimeters)	Weight (pounds)
A	63	160	150
B	67	170.2	160
C	70	177.8	171

Height (inches), weight

Distance from A to B: 10.77

Distance from A to C: 22.14

Distance from B to C: 11.40

Height (centimeters), weight

Distance from A to B: 14.28

Distance from A to C: 27.53

Distance from B to C: 13.37

Important: Always scale your data before training a KNN algorithm on it.

Feature Scaling - Two Methods

1. Minmax Scaler:

$$x_{new} = \frac{x - \min_x}{\max_x - \min_x}$$

2. Standard Scaler:

$$x_{new} = \frac{x - mean_x}{std_x}$$

Strengths and Weaknesses of KNN

- Strengths:
  - Very easy to understand
  - Does not require a lot of adjustments
  - Offers a good baseline
- Weaknesses:
  - For large samples, the performance of KNN degrades
  - Curse of dimensionality (performs poorly on a large number of features)
  - Does not perform well on sparse dataset