



COMP 2211 Exploring Artificial Intelligence  
K-Nearest Neighbor - Population and Sample Standard Deviation  
Dr. Desmond Tsoi

Department of Computer Science & Engineering  
The Hong Kong University of Science and Technology, Hong Kong SAR, China



## Standard Deviation

- Standard deviation is one of the most common ways to measure the spread of values in a dataset.
- There are two different types of standard deviation you can calculate, depending on the type of data you are working with.
  - Population standard deviation
  - Sample standard deviation

## Population Standard Deviation

- We calculate the population standard deviation when the dataset we are working with represents the entire population, i.e. every value that we are interested in.
- The formula to calculate a population standard deviation, denoted as  $\sigma$ , is:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

where

- $\sum$ : Summation
- $x_i$ : The  $i$ th value in the dataset
- $\mu$ : The population mean
- $n$ : The population size

## Sample Standard Deviation

- We calculate the sample standard deviation when the dataset we are working with represents a sample taken from a large population of interest.
- The formula to calculate a sample standard deviation, denoted as  $s$ , is:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

where

- $\sum$ : Summation
- $x_i$ : The  $i$ th value in the dataset
- $\bar{x}$ : The sample mean
- $n$ : The sample size

## Population vs. Sample Standard Deviation

- The difference between the population and the sample standard deviation: When calculating the sample standard deviation, we divided by  $n-1$  instead of  $n$ .
- Because when we calculate the sample standard deviation, we tend to underestimate the true variability in the population. In other words, our estimate of the true population standard is biased.
- To correct this bias, we divide by  $n-1$ . This makes the sample standard deviation an unbiased estimate of the population standard deviation.

That's all!

Any questions?

