standard deviation and variance in detail with an example

ChatGPT

Standard deviation and variance are measures of the spread or dispersion of a set of values. They provide insight into how much individual values in a data set deviate from the mean (average) value.

1. **Variance:**
   - **Definition:** Variance is the average of the squared differences from the mean.
   - **Formula:** If $X_1, X_2, ..., X_n$ are the individual data points, and $\bar{X}$ is the mean, then the variance σ^2 is calculated as:
     $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$
   - **Example:** Consider a data set of exam scores: 75, 80, 85, 90, 95. The mean is $\bar{X} = \frac{75+80+85+90+95}{5} = 85$. The variance is calculated as:
     $\sigma^2 = \frac{(75-85)^2+(80-85)^2+(85-85)^2+(90-85)^2+(95-85)^2}{5}$
     $\sigma^2 = \frac{100+25+0+25+100}{5} = \frac{250}{5} = 50$

2. **Standard Deviation:**
   - **Definition:** Standard deviation is the square root of the variance. It measures the average distance between each data point and the mean.
   - **Formula:** If σ^2 is the variance, then the standard deviation σ is calculated as:
     $\sigma = \sqrt{\sigma^2}$
   - **Example:** Using the variance example above, the standard deviation is:
     $\sigma = \sqrt{50} \approx 7.07$

Interpretation:

- A small standard deviation or variance indicates that the data points tend to be close to the mean.
- A large standard deviation or variance indicates that the data points are spread out over a wider range.

In the example, a standard deviation of approximately 7.07 suggests that the exam scores are somewhat dispersed around the mean sco ↓ 85. If the standard deviation were smaller, it would imply that the scores are more tightly clustered around the mean.