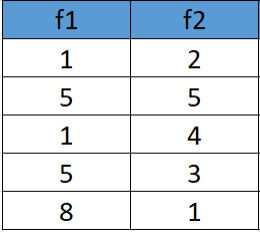
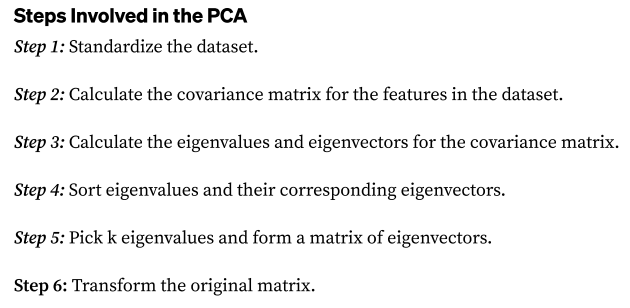


**Fig 1.** An illustration of reducing the dimensions of data

| **Covariance** | **Correlation** |
| --- | --- |
| **Measures solely the linear relationship between variables** | **Measures both the direction and strength of the linear relationship between the variables** |
| **Takes the units of the variables it is measuring** | **Unit-free measurement** |
| **Will be affected by changes in scale** | **Will not be affected by changes in scale** |
| **Limited to two variables** | **Able to be used for several sets of numbers** |

**Consider the following example with 2 features and 5 training examples on k=1 (Reducing dimensions from 2 to 1):**





**Step 1: Standardizing the dataset:**

**First, we calculate the mean and standard deviation for each feature.**

A math equations with numbers and formulas

Description automatically generated with medium confidence

**Standardized Dataset:**

**A number of numbers on a white background

Description automatically generated**

**Step 2: Covariance Matrix:**

**A white background with black text

Description automatically generated**

**A number of numbers on a white background

Description automatically generated**

**A math equations and numbers

Description automatically generated with medium confidence**

**Step 3,4,5: Calculate eigenvectors and eigenvalues of the covariance matrix, sort them and choose k features based on selection number k:**

**A math problem with numbers and equations

Description automatically generated**

**A math problem with numbers and lines

Description automatically generated**

**NOTE: THERE ARE TWO EIGENVALUES, SINCE K = 1 WE CHOOSE THE HIGHEST EIGENVALUE (1.9965)**

**Only**

**A white sheet with numbers and lines

Description automatically generated**

**A white paper with black text and numbers

Description automatically generated-\***

**A math problem with numbers

Description automatically generated**