Mean, Variance, Correlation and Covariance

First a review of the **1-dimensional** case. Later the process is repeated for the multivariate case.

Sample Mean

We have a measurement X of `n` samples $x_1, \ldots, x_n$. The mean / mean value is the defined as the average:

$$E(X) = \frac{1}{n}\sum_{j=1}^{n} x_j$$

If we choose to arrage samples into a column vector $\mathbf{x}$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

we can write

$$E(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n} x_j$$

The mean value of the squared elements $x_1^2, \ldots, x_n^2$ is computed from

$$E(X^2) = \frac{1}{n}\sum_{j=1}^{n} x_j^2$$

or using data $\mathbf{x}$ arranged in a column vector:

$$E(X^2) = \frac{1}{n}\mathbf{x}^T \cdot \mathbf{x}$$

Sample Variance

$$E((X-E(X))^2 = \frac{1}{n}\sum_{j=1}^{n}\left(x_j - E(X)\right)^2$$

$$= \frac{1}{n}\sum_{j=1}^{n} x_j^2 - 2\frac{1}{n}\sum_{j=1}^{n} x_j \cdot E(X) + \frac{1}{n}\sum_{j=1}^{n} E(X)^2$$

$$\rightarrow$$

$$E((X-E(X))^2 = E(X^2) - E(X)^2$$

```
In [ ]:
```

Mean / Expectation of a vector

From `N` measurements each measurement produces a *data point* of `K` items.

We assume that the `j` th measurement yields a data point that is represented as a row vector $\mathbf{d}_j^T$

$$\mathbf{d}_j^T = \begin{bmatrix} d_{j,\,1} & \cdots & d_{j,\,K} \end{bmatrix}$$

We will arrange the `N` measurements into a *data matrix* $\mathbf{D} :\in \mathrm{R}^{N \times K}$:

$$\mathbf{D} = \begin{bmatrix} d_{1,\,1} & d_{1,\,2} & \cdots & d_{1,\,K} \\ \vdots & \vdots & \vdots & \vdots \\ d_{j,\,1} & d_{j,\,2} & \cdots & d_{j,\,K} \\ \vdots & \vdots & \vdots & \vdots \\ d_{N,\,1} & d_{N,\,2} & \cdots & d_{N,\,K} \end{bmatrix}$$

Each row of $\mathbf{D}$ represents a single measurement of `K` items (eg.: temperature, time, voltage, speed, ...).

The `i-th` column vector $\mathbf{d}_{j:\,,\,i}$ of $\mathbf{D}$ contains all measurements of the `i-th` measurement item (eg: temperature).

Hence the mean value of the `i-th` measurement item is just the mean value of the elements of column vector $\mathbf{d}_{j:\,,\,i}$:

$$E(\mathbf{d}_{j:\,,\,i}) = \frac{1}{N}\sum_{j=1}^{N} d_{j,i}$$

In some cases it is necessary to remove the mean value of each data column from its data column to obtain the *centered* data matrix $\bar{\mathbf{D}}$

$$\bar{\mathbf{D}} = \begin{bmatrix} \left(d_{1,1} - E(\mathbf{d}_{j:,1})\right) & \left(d_{1,2} - E(\mathbf{d}_{j:,2})\right) & \cdots & \left(d_{1,K} - E(\mathbf{d}_{j:,K})\right) \\ \vdots & \vdots & \vdots & \vdots \\ \left(d_{j,1} - E(\mathbf{d}_{j:,1})\right) & \left(d_{j,2} - E(\mathbf{d}_{j:,2})\right) & \cdots & \left(d_{j,K} - E(\mathbf{d}_{j:,K})\right) \\ \vdots & \vdots & \vdots & \\ \left(d_{N,1} - E(\mathbf{d}_{j:,1})\right) & \left(d_{N,2} - E(\mathbf{d}_{j:,2})\right) & \cdots & \left(d_{N,K} - E(\mathbf{d}_{j:,K})\right) \end{bmatrix}$$

Random Vector

$\mathbf{x}: \ \in \mathrm{R}^K$.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_K \end{bmatrix}$$

Now we assume there are `N` realisation of such a random vector. We denote the `j-th` realisation by $\mathbf{x}_j$ and its elements / components by:

$$\mathbf{x}_j = \begin{bmatrix} x_{1,j} \\ \vdots \\ x_{i,j} \\ \vdots \\ x_{K,j} \end{bmatrix}$$

We define the expectation $E(\mathbf{x}$ of these `N` random vector element-wise like this:

$$E(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_j = \begin{bmatrix} \frac{1}{N}\sum_{j=1}^{N}x_{1,j} \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N}x_{i,j} \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N}x_{K,j} \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_i \\ \vdots \\ \bar{x}_K \end{bmatrix} = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_i) \\ \vdots \\ E(x_K) \end{bmatrix}$$

Now we consider the matrix equation

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x} + \mathbf{b}:\ \mathbf{A} \in \mathrm{R}^{L \times K};\ \mathbf{x} \in \mathrm{R}^{K};\ \mathbf{b} \in \mathrm{R}^{L};\ \mathbf{y} \in \mathrm{R}^{L}$$

If we apply the data vectors $\mathbf{x}_j$: $j = 1, \ldots, N$ to this matrix equation we get transformed data vectors $\mathbf{y}_j$: $j = 1, \ldots, N$

We want to compute the expection $E(\mathbf{y}$:

$$E(\mathbf{y}) = \frac{1}{N}\sum_{j=1}^{N}\mathbf{y}_j + \mathbf{b} = \begin{bmatrix} \frac{1}{N}\sum_{j=1}^{N} y_{1,j} \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} y_{i,j} \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} y_{L,j} \end{bmatrix} + \mathbf{b}$$

$$= \begin{bmatrix} \frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{K} a_{(1,\,k)} \cdot x_{(k,j)} \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} \end{bmatrix} + \mathbf{b} = \begin{bmatrix} \sum_{k=1}^{K} a_{(1,\,k)} \cdot \left(\frac{1}{N}\sum_{j=1}^{N} x_{(k,j)}\right) \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} \end{bmatrix} + \mathbf{b}$$

$$= \begin{bmatrix} \sum_{k=1}^{K} a_{(1,\,k)} \cdot E(\mathbf{x}) \\ \vdots \\ \sum_{k=1}^{K} a_{(i,\,k)} \cdot E(\mathbf{x}) \\ \vdots \\ \sum_{k=1}^{K} a_{(L,\,k)} \cdot E(\mathbf{x}) \end{bmatrix} + \mathbf{b} = \mathbf{A} \cdot E(\mathbf{x}) + \mathbf{b}$$

Important here is the fact that compute the expectation vector $E(\mathbf{y})$ there is no need to compute `N` transformed data vectors $\mathbf{y}_j$. The expectation vector $E(\mathbf{x})$ can be directly applied to the matrix equation. Thus quite a number of matrix/vector multiplication are saved.

Variance

We examine again *data matrix* $\mathbf{D} :\in \mathrm{R}^{N \times K}$ with `N` measurements. Each measurement has `K` items. Thus each row of $\mathbf{D}$ represents a single measurement of `K` items (eg.: temperature, time, voltage, speed, ...).

The `j-th` measurement has `K` measured items. These items are arranged into a column vector denoted $\mathbf{d}_j$.

$$\mathbf{d}_j = \begin{bmatrix} d_1[j] \\ \vdots \\ d_k[j] \\ \vdots \\ d_K[j] \end{bmatrix} = \begin{bmatrix} d_{1,j} \\ \vdots \\ d_{kj} \\ \vdots \\ d_{Kj} \end{bmatrix}$$

$d_k[j] = d_{kj}$ denotes the `j-th` measurement of the `k-th` item.

Now we define a vector $\mathbf{w} : \in \mathrm{R}^K$. This vector shall be used to compute a weighted addition of each measurement. For each measurement we compute the dot product $\mathbf{w}^T\mathbf{d}_j : j = 1, \ldots, N$.

For each measurement we get a scalar s_j:

$$s_j = \mathbf{w}^T\mathbf{d}_j$$

The average value of these `N` s_j is denoted $E(s)$ and computed from

$$E(s) = \frac{1}{N}\sum_{j=1}^{N} s_j = \mathbf{w}^T \cdot \underbrace{\frac{1}{N}\sum_{j=1}^{N} \mathbf{d}_j}_{E(\mathbf{d})} = \mathbf{w}^T \cdot E(\mathbf{d})$$

$E(\mathbf{d})$ denotes the element wise expectation of data items. Vector $E(\mathbf{d}) : \in \mathrm{R}^K$ can be expressed like this:

$$E(\mathbf{d}) = \begin{bmatrix} \frac{1}{N}\sum_{j=1}^{N} d_1[j] \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} d_k[j] \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} d_K[j] \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_{j=1}^{N} d_{1,j} \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} d_{k,j} \\ \vdots \\ \frac{1}{N}\sum_{j=1}^{N} d_{K,j} \end{bmatrix} = \begin{bmatrix} E(d_1) \\ \vdots \\ E(d_k) \\ \vdots \\ E(d_K) \end{bmatrix}$$

$E(d_k) = \frac{1}{N}\sum_{j=1}^{N} d_k[j] = \frac{1}{N}\sum_{j=1}^{N} d_{k,j}$ is the mean value / expected value of the `k-th` measurement item.

---

centered data set

$$\mathbf{c}_j = \mathbf{d}_j - E(\mathbf{d}) = \begin{bmatrix} d_1[j] \\ \vdots \\ d_k[j] \\ \vdots \\ d_K[j] \end{bmatrix} - \begin{bmatrix} E(d_1) \\ \vdots \\ E(d_k) \\ \vdots \\ E(d_K) \end{bmatrix} = \begin{bmatrix} d_{1,j} \\ \vdots \\ d_{kj} \\ \vdots \\ d_{Kj} \end{bmatrix} - \begin{bmatrix} E(d_1) \\ \vdots \\ E(d_k) \\ \vdots \\ E(d_K) \end{bmatrix}$$

$$g_j = \mathbf{w}^T\mathbf{c}_j = \mathbf{w}^T \cdot \left(\mathbf{d}_j - E(\mathbf{d})\right) = s_j - \mathbf{w}^T \cdot E(\mathbf{d})$$

The squared value g_j^2

$$g_j^2 = \left(\mathbf{w}^T\mathbf{c}_j\right)^2 = \left(\mathbf{w}^T\mathbf{c}_j\right) \cdot \left(\mathbf{c}_j^T\mathbf{w}\right) = \mathbf{w}^T \cdot \left(\mathbf{c}_j \cdot \mathbf{c}_j^T\right) \cdot \mathbf{w}$$

Defining the square matrix $\mathbf{C}_j: \ \in \mathrm{R}^{K \times K}$ by:

$$\mathbf{C}_j = \mathbf{c}_j \cdot \mathbf{c}_j^T = \left[\begin{matrix} \left(d_{1,j} - E(d_1)\right) \cdot \left(d_{1,j} - E(d_1)\right) & \cdots & \left(d_{1,j} - E(d_1)\right) \cdot \left(d_{k,j} - E(d_k)\right) & \cdots & \left(d_{1,}\right. \\ \vdots & \vdots & \cdots & \vdots & \\ \left(d_{k,j} - E(d_k)\right) \cdot \left(d_{1,j} - E(d_1)\right) & \cdots & \left(d_{k,j} - E(d_k)\right) \cdot \left(d_{k,j} - E(d_k)\right) & \cdots & \left(d_{k,}\right. \\ \vdots & \vdots & \cdots & \vdots & \\ \left(d_{K,j} - E(d_K)\right) \cdot \left(d_{1,j} - E(d_1)\right) & \cdots & \left(d_{K,j} - E(d_K)\right) \cdot \left(d_{k,j} - E(d_k)\right) & \cdots & \left(d_{K,}\right. \end{matrix} \right.$$

With the defintion of $\mathbf{C}_j$ we are able to write g_j^2 as:

$$g_j^2 = \mathbf{w}^T \cdot \mathbf{C}_j \cdot \mathbf{w}$$

And the expectation as

$$E(g^2) = \mathbf{w}^T \cdot \left(\frac{1}{N}\sum_{j=1}^{N}\mathbf{C}_j\right) \cdot \mathbf{w}$$

$$\mathbf{C} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{C}_j$$

$$\mathbf{C} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{c}_j \cdot \mathbf{c}_j^T = \begin{bmatrix} \frac{1}{N}\sum_{j=1}^{N}\Big(d_{1,j} - E(d_1)\Big) \cdot \Big(d_{1,j} - E(d_1)\Big) & \cdots & \frac{1}{N}\sum_{j=1}^{N}\Big(d_{1,j} - E(d_1)\Big) \cdot \Big(d_{k,j} - E \\ \vdots & \vdots & \ldots \\ \frac{1}{N}\sum_{j=1}^{N}\Big(d_{k,j} - E(d_k)\Big) \cdot \Big(d_{1,j} - E(d_1)\Big) & \cdots & \frac{1}{N}\sum_{j=1}^{N}\Big(d_{k,j} - E(d_k)\Big) \cdot \Big(d_{k,j} - E \\ \vdots & \vdots & \ldots \\ \frac{1}{N}\sum_{j=1}^{N}\Big(d_{K,j} - E(d_K)\Big) \cdot \Big(d_{1,j} - E(d_1)\Big) & \cdots & \frac{1}{N}\sum_{j=1}^{N}\Big(d_{K,j} - E(d_K)\Big) \cdot \Big(d_{k,j} - E \end{bmatrix}$$

The elements of matrix $\mathbf{C} :\in \mathrm{R}^{K \times K}$ are denoted $v_{l,\,m}$: $l = 1, \ldots, K;\; m = 1, \ldots, K.$

$$v_{l,\,m} = \frac{1}{N}\sum_{j=1}^{N}\Big(d_{l,j} - E(d_l)\Big) \cdot \Big(d_{m,j} - E(d_m)\Big)$$

case: $\boldsymbol{l = m}$ (diagonal elements of $\mathbf{C}$)

$$v_{l,\,l} = \frac{1}{N}\sum_{j=1}^{N}\Big(d_{l,j} - E(d_l)\Big)^2$$

$$= \frac{1}{N}\sum_{j=1}^{N} d_{l,j}^2 - E(d_l)^2 = E(d_l^2) - E(d_l)^2 = Variance(d_l) = Var(d_l)$$

case: $\boldsymbol{l \neq m}$ (off-diagonal elements of $\mathbf{C}$)

$$\begin{aligned} v_{l,\,m} &= \frac{1}{N}\sum_{j=1}^{N} d_{l,j} \cdot d_{m,j} - E(d_m) \cdot \frac{1}{N}\sum_{j=1}^{N} d_{l,j} - E(d_l) \cdot \frac{1}{N}\sum_{j=1}^{N} d_{m,j} + E(d_l) \cdot E(d_m) \\ &= E(d_{l,j} \cdot d_{m,j}) - 2 \cdot E(d_l) \cdot E(d_m) + E(d_l) \cdot E(d_m) \\ &= E(d_l \cdot d_m) - E(d_l) \cdot E(d_m) = Covariance(d_l, d_m) = Cov(d_l, d_m) \end{aligned}$$

In [ ]:

In [ ]: