

# Data Exploration

11.02.2020

HW 01

$$x_i = \begin{bmatrix} x_i^1 \\ x_i^2 \\ \vdots \\ x_i^m \end{bmatrix} \quad i = \overline{1 \dots n}$$

Sample mean vector

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i - \text{[number]} - \text{vector},$$

Sample covariance matrix

$$V = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$$

$T$  - transpose

$$\begin{bmatrix} x_i \\ \vdots \\ x_i \end{bmatrix} \cdot \begin{bmatrix} x_i^1 & \dots & x_i^m \end{bmatrix}$$

What canonical form of information would you suggest to represent the seq.  $x_i$  in order to compute the Sample mean Vector and the sample covariance matrix?

Observations for Sample mean vector

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i;$$

The result  $\bar{x}$  depends on count of vectors, and the sum of received vectors. So, for this SMV, we need to store  $n$ , and the sum

$$\begin{bmatrix} x_1 \\ \vdots \\ x_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ \vdots \\ x_2 \end{bmatrix}$$

$(n, s)$