Multivariate Statistics

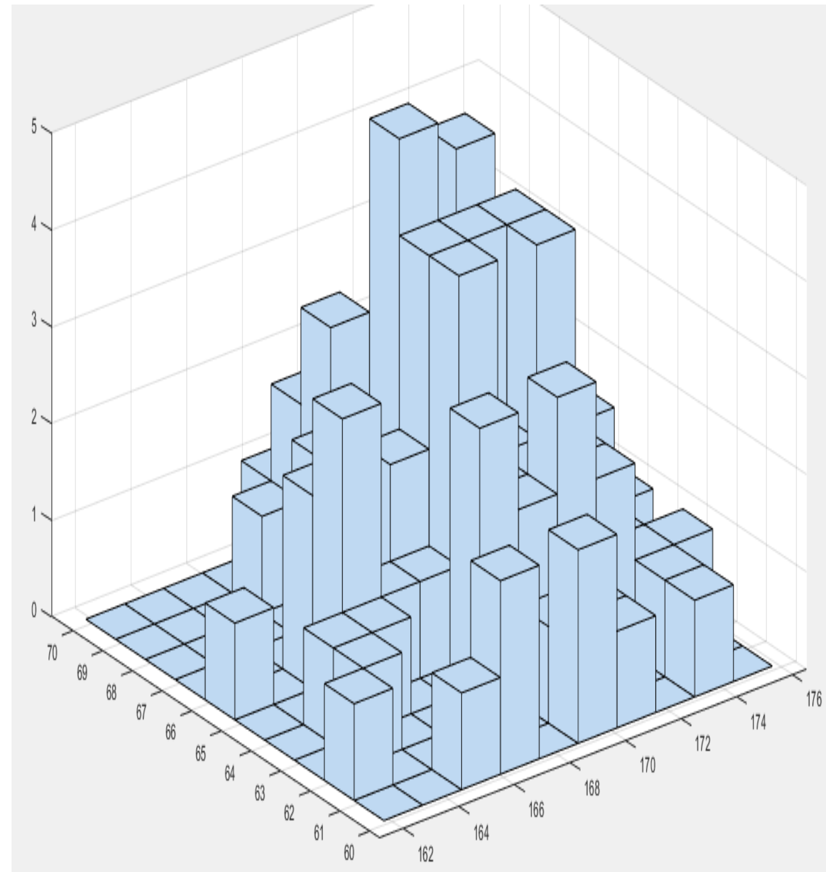
Dr. Pritam Anand.
Assistant Professor,
DA-IICT, Gandhinagar.

[Weights in Kg , Heights in Cm] (X,Y)

[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 148]
[102, 780]	[83, 163.1]	[97,182]
[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 188]
[90, 168]	[83, 163.1]	[87,182]
[55, 161.8]	[84, 162.6]	[62, 181.2]
[88,172]	[90, 157.5]	[69,170]
[57, 169]	[62, 149.7]	[63 , 142]
[102, 178]	[83, 173.1]	[97,181]
[53, 168]	[84, 162.6]	[62, 179.2]
[88,170]	[90, 177.5]	[105,174]
[57, 175]	[62, 169.7]	[53 , 146]
[102, 168]	[83, 168.1]	[97,180]

[Weights in Kg , Heights in Cm] (X,Y)

[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 148]
[102, 780]	[83, 163.1]	[97,182]
[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 188]
[90, 168]	[83, 163.1]	[87,182]
[55, 161.8]	[84, 162.6]	[62, 181.2]
[88,172]	[90, 157.5]	[69,170]
[57, 169]	[62, 149.7]	[63 , 142]
[102, 178]	[83, 173.1]	[97,181]
[53, 168]	[84, 162.6]	[62, 179.2]
[88,170]	[90, 177.5]	[105,174]
[57, 175]	[62, 169.7]	[53 , 146]
[102, 168]	[83, 168.1]	[97,180]



$\frac{X^T y}{\|y\|}$

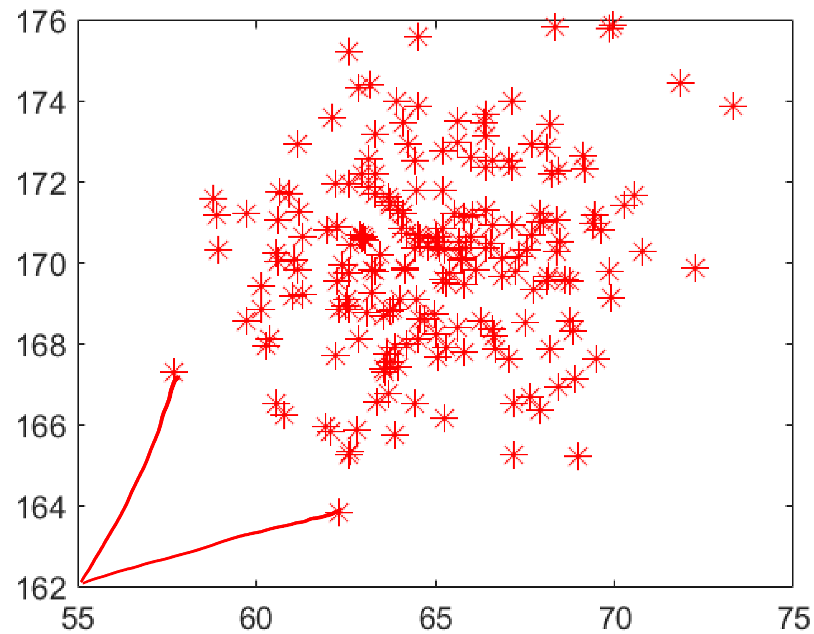
[Weights in Kg , Heights in Cm] (X,Y)

[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 148]
[102, 780]	[83, 163.1]	[97,182]
[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 188]
[90, 168]	[83, 163.1]	[87,182]
[55, 161.8]	[84, 162.6]	[62, 181.2]
[88,172]	[90, 157.5]	[69,170]
[57, 169]	[62, 149.7]	[63 , 142]
[102, 178]	[83, 173.1]	[97,181]
[53, 168]	[84, 162.6]	[62, 179.2]
[88,170]	[90, 177.5]	[105,174]
[57, 175]	[62, 169.7]	[53 , 146]
[102, 168]	[83, 168.1]	[97,180]

$\frac{1}{n}\sum |x_i - y_i|$

R^2 (X_1, X_2) R^2

$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$, $\sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}$



$u = \left(\frac{1}{n}\sum x_i, \frac{1}{n}\sum y_i\right)$

[Weights in Kg , Heights in Cm] (X,Y)

[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 148]
[102, 780]	[83, 163.1]	[97,182]
[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 188]
[90, 168]	[83, 163.1]	[87,182]
[55, 161.8]	[84, 162.6]	[62, 181.2]
[88,172]	[90, 157.5]	[69,170]
[57, 169]	[62, 149.7]	[63 , 142]
[102, 178]	[83, 173.1]	[97,181]
[53, 168]	[84, 162.6]	[62, 179.2]
[88,170]	[90, 177.5]	[105,174]
[57, 175]	[62, 169.7]	[53 , 146]
[102, 168]	[83, 168.1]	[97,180]

$$\sum_{i=1}^{n} (x_i - c)^T (x_i - c)$$

Mean ??

$$\sum_{i=1}^{n} \|x_i - c\|^2 \text{ is } \underline{\underline{\text{minimum}}}$$

Norm of a vector

$x \in R^n$

$$\|x\|_2 = \left| \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \right|$$

$$\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$$

$$\|x\|_p = (x_1^p + x_2^p + \dots \quad x^p)^{1/p}$$

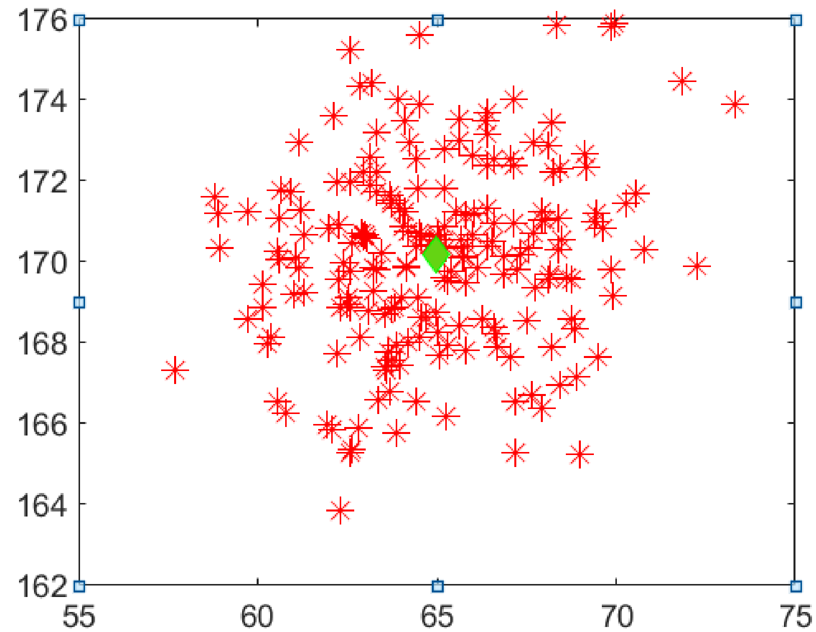
$$\|x\|_\infty = \max(x_1, x_2, \dots x_n)$$

$$\frac{x^T y}{\|y\|} =$$

[Weights in Kg , Heights in Cm] (X,Y)

[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 148]
[102, 780]	[83, 163.1]	[97,182]
[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 188]
[90, 168]	[83, 163.1]	[87,182]
[55, 161.8]	[84, 162.6]	[62, 181.2]
[88,172]	[90, 157.5]	[69,170]
[57, 169]	[62, 149.7]	[63 , 142]
[102, 178]	[83, 173.1]	[97,181]
[53, 168]	[84, 162.6]	[62, 179.2]
[88,170]	[90, 177.5]	[105,174]
[57, 175]	[62, 169.7]	[53 , 146]
[102, 168]	[83, 168.1]	[97,180]

$$||X||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$



$$||X||_1 = \sum_{i=1}^{n} |x_i|$$

[Weights in Kg , Heights in Cm] (X,Y)

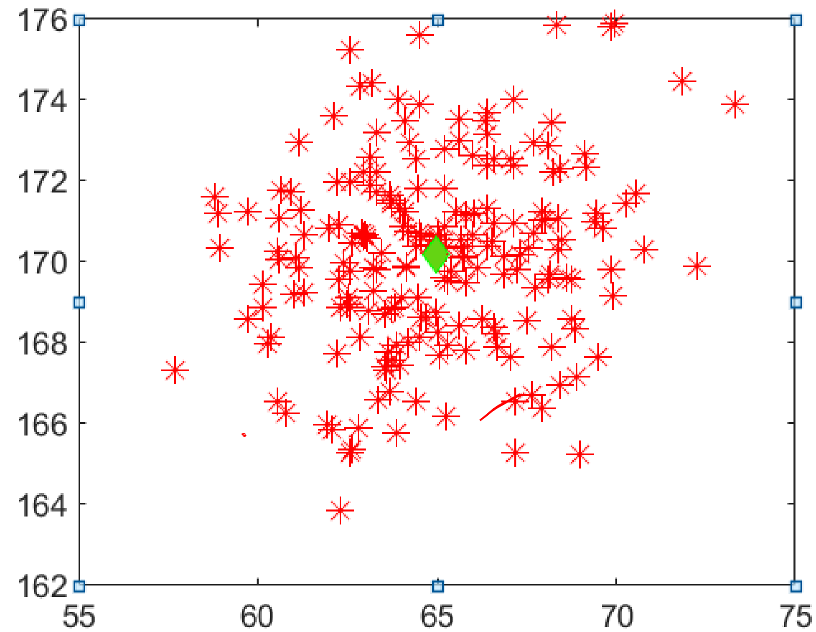
[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 148]
[102, 780]	[83, 163.1]	[97,182]
[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 188]
[90, 168]	[83, 163.1]	[87,182]
[55, 161.8]	[84, 162.6]	[62, 181.2]
[88,172]	[90, 157.5]	[69,170]
[57, 169]	[62, 149.7]	[63 , 142]
[102, 178]	[83, 173.1]	[97,181]
[53, 168]	[84, 162.6]	[62, 179.2]
[88,170]	[90, 177.5]	[105,174]
[57, 175]	[62, 169.7]	[53 , 146]
[102, 168]	[83, 168.1]	[97,180]

Median ??

Project data

$d \in R^n$

[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 148]
[102, 780]	[83, 163.1]	[97,182]
[55, 161.8]	[84, 162.6]	[62, 159.2]
[88,172]	[90, 157.5]	[105,172]
[57, 169]	[62, 149.7]	[53 , 188]
[90, 168]	[83, 163.1]	[87,182]
[55, 161.8]	[84, 162.6]	[62, 181.2]
[88,172]	[90, 157.5]	[69,170]
[57, 169]	[62, 149.7]	[63 , 142]
[102, 178]	[83, 173.1]	[97,181]
[53, 168]	[84, 162.6]	[62, 179.2]
[88,170]	[90, 177.5]	[105,174]
[57, 175]	[62, 169.7]	[53 , 146]
[102, 168]	[83, 168.1]	[97,180]



$u = \frac{1}{N} \sum_{i=1}^{N} x_i$

$d^T x_i$

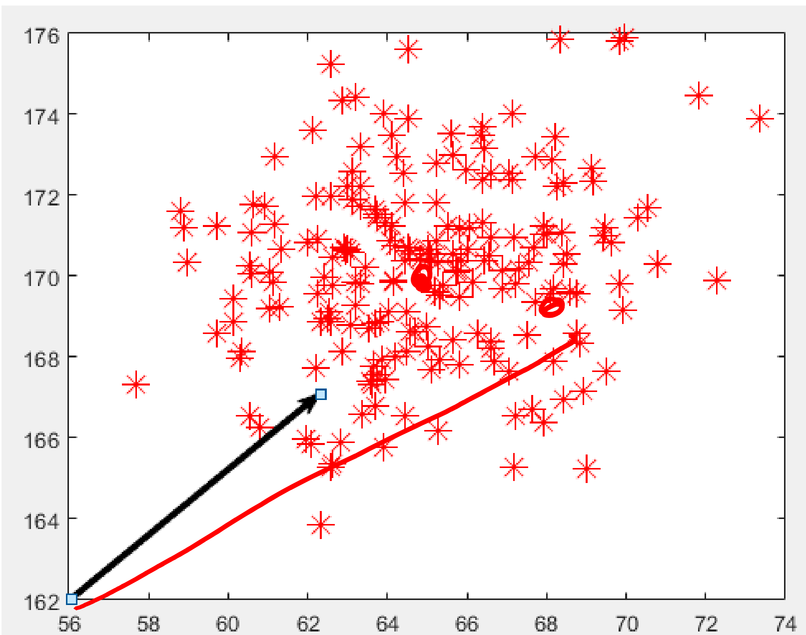
$d^T u$

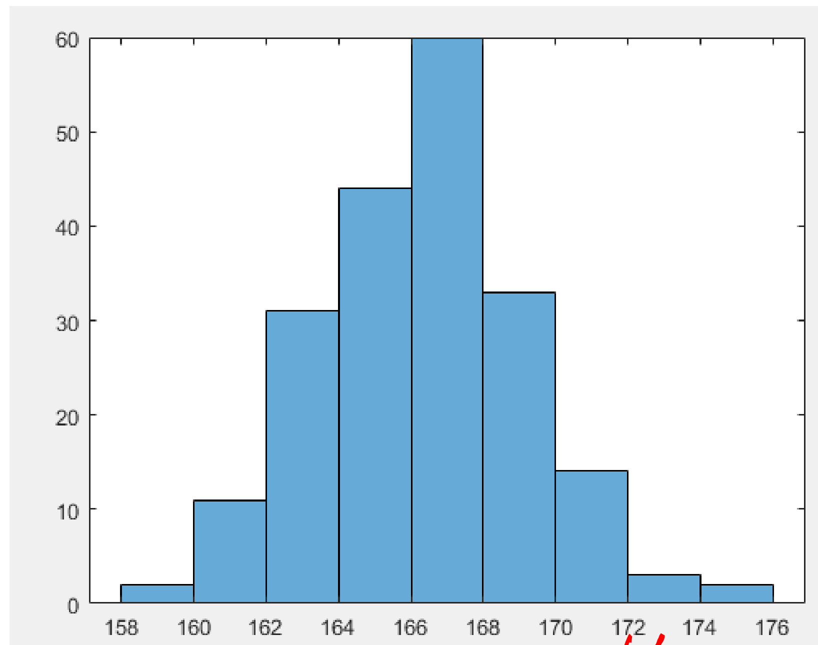
Projected data

$$\frac{1}{N}\sum_{i=1}^{n} d^T x_i$$

$$d^T \frac{1}{N}\sum_{i=1}^{N} x_i$$

x_i $\quad x$





$$\frac{1}{N}\left(d^T x_1 + d^T x_2 + \cdots d^T x_N\right) = d^T \frac{1}{N}\overbrace{\left(x_1 + x_2 + \cdots x_N\right)}^{u}$$

X_1 $\begin{bmatrix} x_{11} \\ x_{12} \\ - \\ x_{1n} \end{bmatrix}$ $d^T x_1$

$$u = \frac{1}{l} \sum X_i$$ - mean

$d \in R^n$

$\|d\|_2 = 1$

$-$

X_2 $\begin{bmatrix} x_{21} \\ x_{22} \\ - \\ x_{2n} \end{bmatrix}$ $d^T x_2$

$-$

$-$

projected mean $= \frac{1}{l} \sum d^T x_i$

$$= d^T \left(\frac{1}{l} \sum_{i=1}^{l} x_i \right) = d^T u$$

X_l $\begin{bmatrix} x_{l1} \\ x_{l2} \\ - \\ x_{ln} \end{bmatrix}$ $d^T x_l$

$$d^T (x + y + z) = d^T x + d^T y + d^T z$$

$d^T x_1 \quad d^T x_2 \quad \cdots \quad d^T x_N$

$\downarrow d^T u$

A

$x^T \cancel{A} A$

$x^T A x \geq 0$

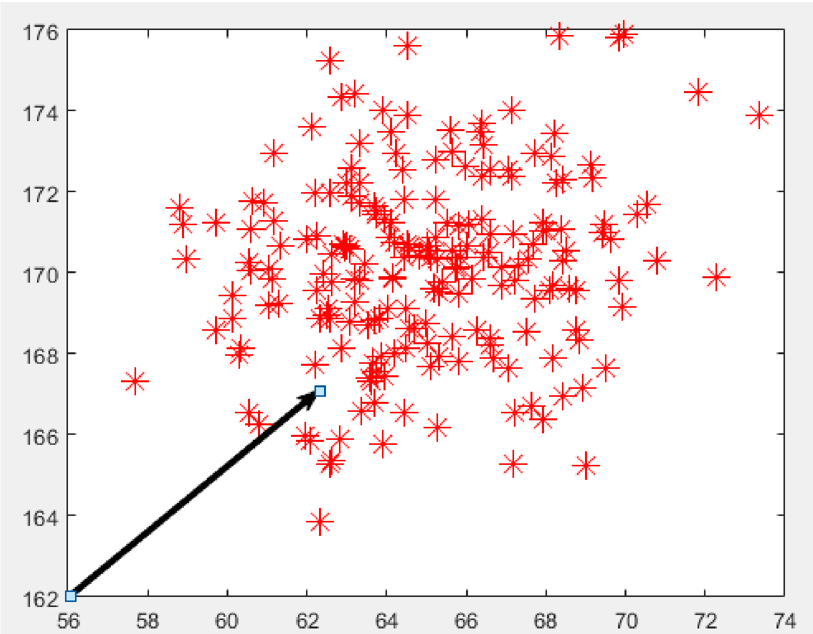
$$\frac{1}{N-1} \sum_{i=1}^{N} \left(d^T x_i - d^T u\right)^2$$

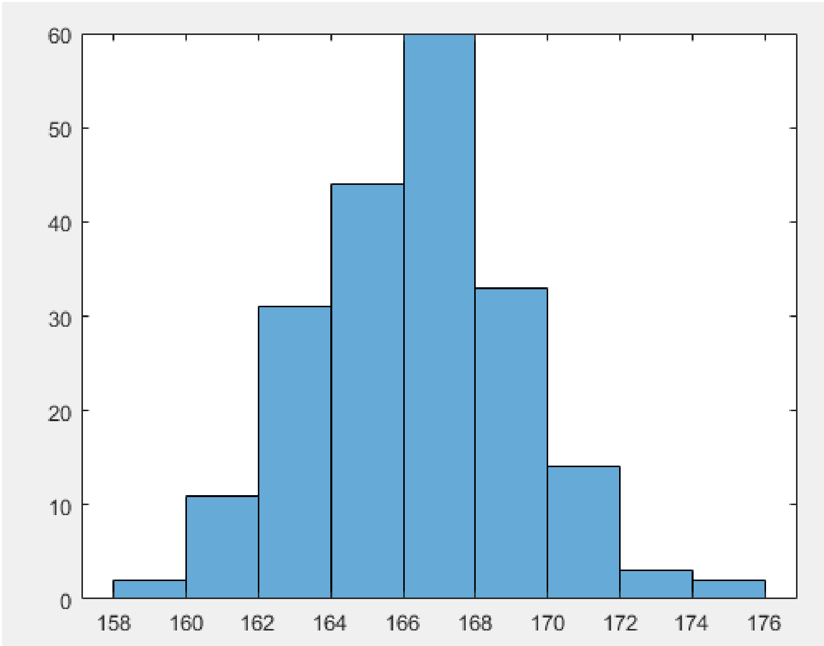
$$\frac{1}{N-1} \sum_{i=1}^{N} \left(d^T (x_i - u)\right)^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(\left(d^T (x_i - u)\right)\left((x_i - u)^T d\right)\right)$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} d^T \left[(x_i - u)(x_i - u)^T\right] d$$

$$d^T \Sigma d \geq 0 = d^T \underbrace{\left(\frac{1}{N-1} \sum (x_i - u)(x_i - u)^T\right)}_{\Sigma} d$$

Mean of Projected data





X_1

X_2

X_N

$$\begin{bmatrix} x_{11}, & x_{21}, & \dots & x_{N1} \\ x_{12} & x_{22}, & \dots & x_{N2} \\ \\ x_{1n} & x_{2n}, & \dots & x_{Nn} \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \\ u_n \end{bmatrix}$$

$$\frac{1}{N-1} \sum_{i=1}^{N} (X_i - u)(X_i - u)^T \qquad n \times n$$

$$\frac{1}{N-1} \begin{bmatrix} x_{11} - u_1 \\ x_{12} - u_2 \\ \\ x_{1n} - u_n \end{bmatrix} [x_{11} - u_1, \; x_{12} - u_2 \; . \; x_{1n} - u_n] + \begin{bmatrix} x_{21} - u_1 \\ x_{22} - u_2 \\ \\ x_{2n} - u_n \end{bmatrix} [\dots$$

$$\dots x_{21} - u_1, \; x_{22} - u_2, \; \dots \; x_{2n} - u_n] + \dots$$

$$\frac{1}{N-1}\begin{bmatrix} \underbrace{(x_{11}-u_1)^2+(x_{21}-u_1)^2,\ldots(x_{n1}-u_n)^2}_{\sigma_1^2} & \ldots & \sigma_{12} & ,\ldots & \sigma_{1n} \\ \sigma_{21} , & & \sigma_2^2 , & \ldots & \sigma_{2n} \\ \sigma_{n1} & & \sigma_{n2} & ,\ldots & \sigma_n^2 \end{bmatrix}$$

$$\left[(x_{11}-\mu_1)(x_{12}-\mu_2) + (x_{21}-\mu_1)(x_{22}-\mu_2) + \dots + (x_{N1}-\mu_1)(x_{N2}-\mu_2) \right]$$

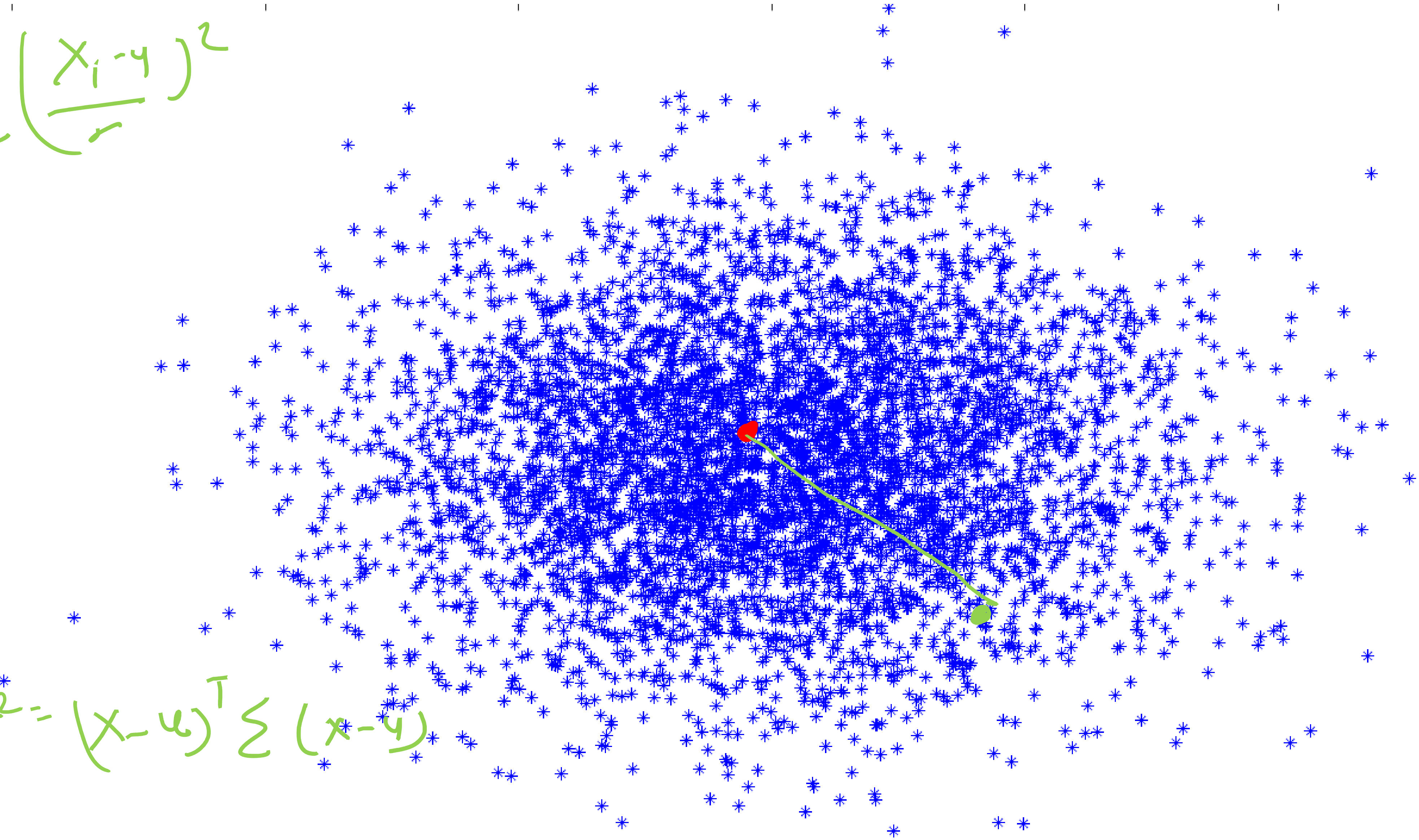
u_1 $\quad$ u_2

x_{11} $\quad$ x_{21}

x_{21} $\quad$ x_{21}

x_{N1} $\quad$ x_{N1}

$$e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$



$$\sigma^2 = (x-\mu)^T \Sigma (x-\mu)$$

Variance of Projected data