Principal Component Analysis

Dimensionality Reduction

Two Popular Methods of Dimension Reduction

- Principal Component Analysis (PCA)

Principal Component Analysis is a methods of dimensionality reduction/feature extraction that transform the data from a d-dimensional space to another coordinate system of k - dimensional space where k<=d.

- Latent Semantic Analysis (LSA)

Latent Semantic Analysis is also another methods of dimensionality reduction originally applied for topic modelling in text corpus. In recent time, it has also been applied to various domains.

Principal Component Analysis

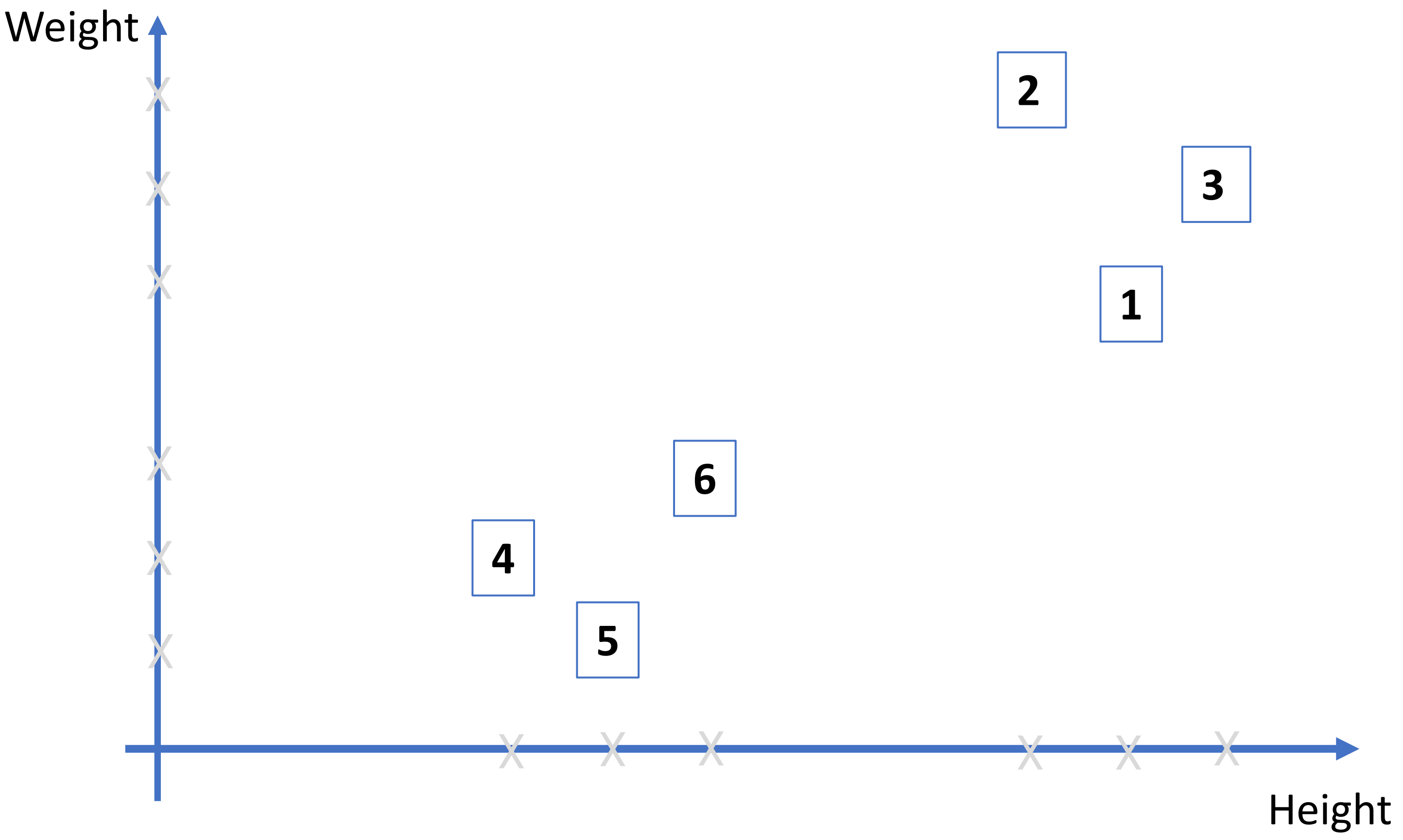
	Height	Weight	#Wheel	CC
1	6	500	4	899
2	5.5	600	4	1000
3	6.5	550	4	800
4	3	200	2	99
5	3.5	150	2	125
6	4	250	2	100

Visualization of the Data with Single Feature

	Height	Weight	#Wheel	CC
1	6	500	4	899
2	5.5	600	4	1000
3	6.5	550	4	800
4	3	200	2	99
5	3.5	150	2	125
6	4	250	2	100

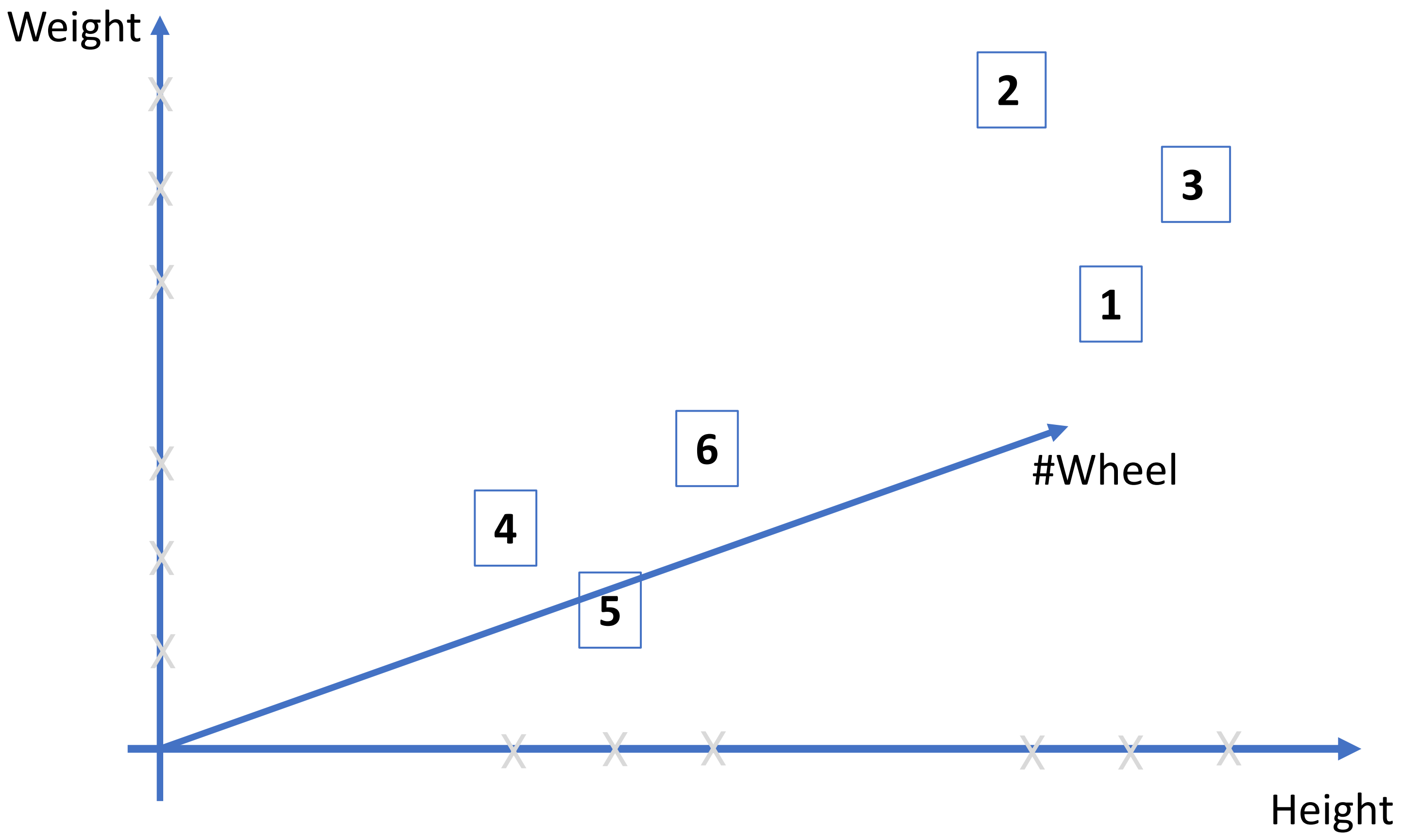
Visualization of the Data with two Features

	Height	Weight	#Wheel	CC
1	6	500	4	899
2	5.5	600	4	1000
3	6.5	550	4	800
4	3	200	2	99
5	3.5	150	2	125
6	4	250	2	100



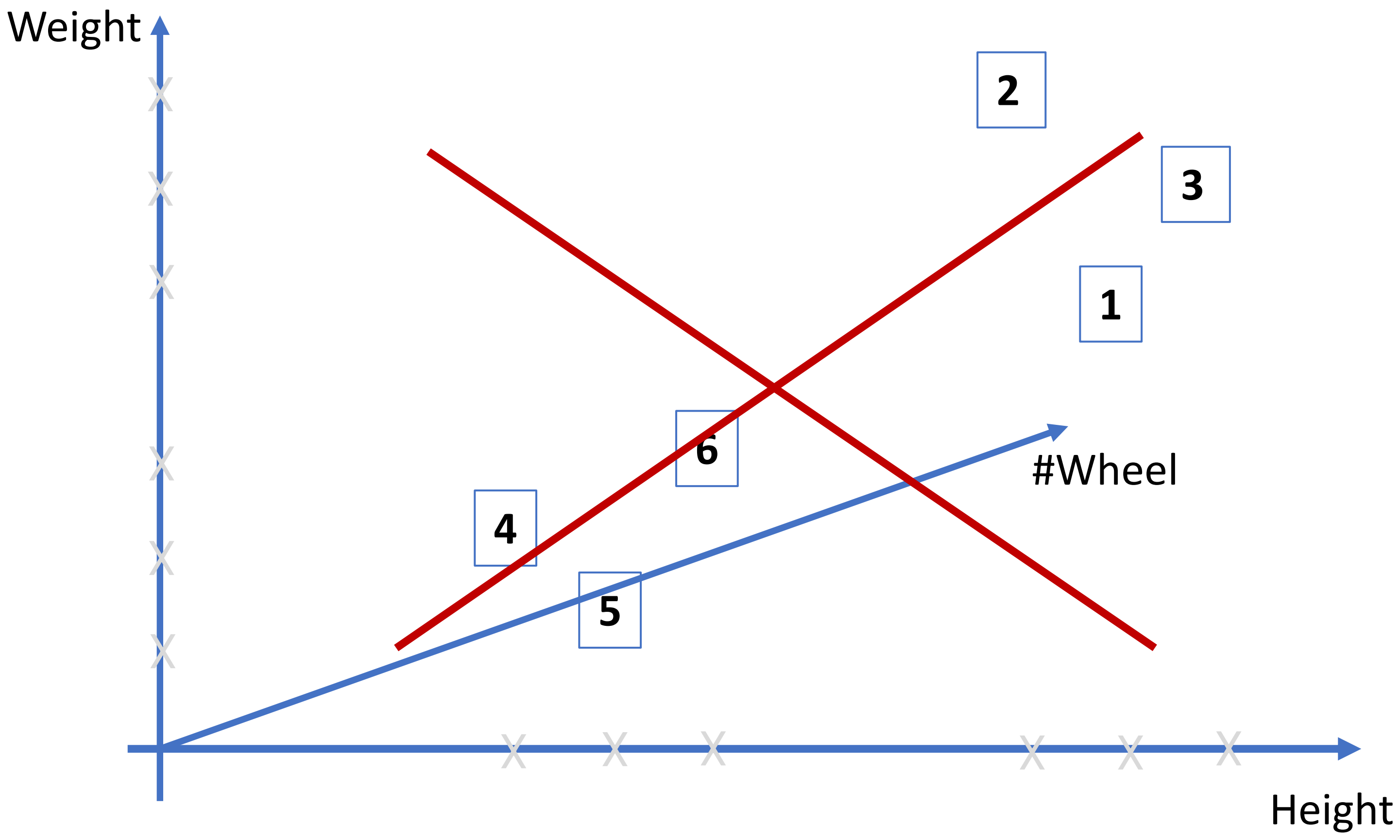
Visualization of the Data with Three Features

	Height	Weight	#Wheel	CC
1	6	500	4	899
2	5.5	600	4	1000
3	6.5	550	4	800
4	3	200	2	99
5	3.5	150	2	125
6	4	250	2	100



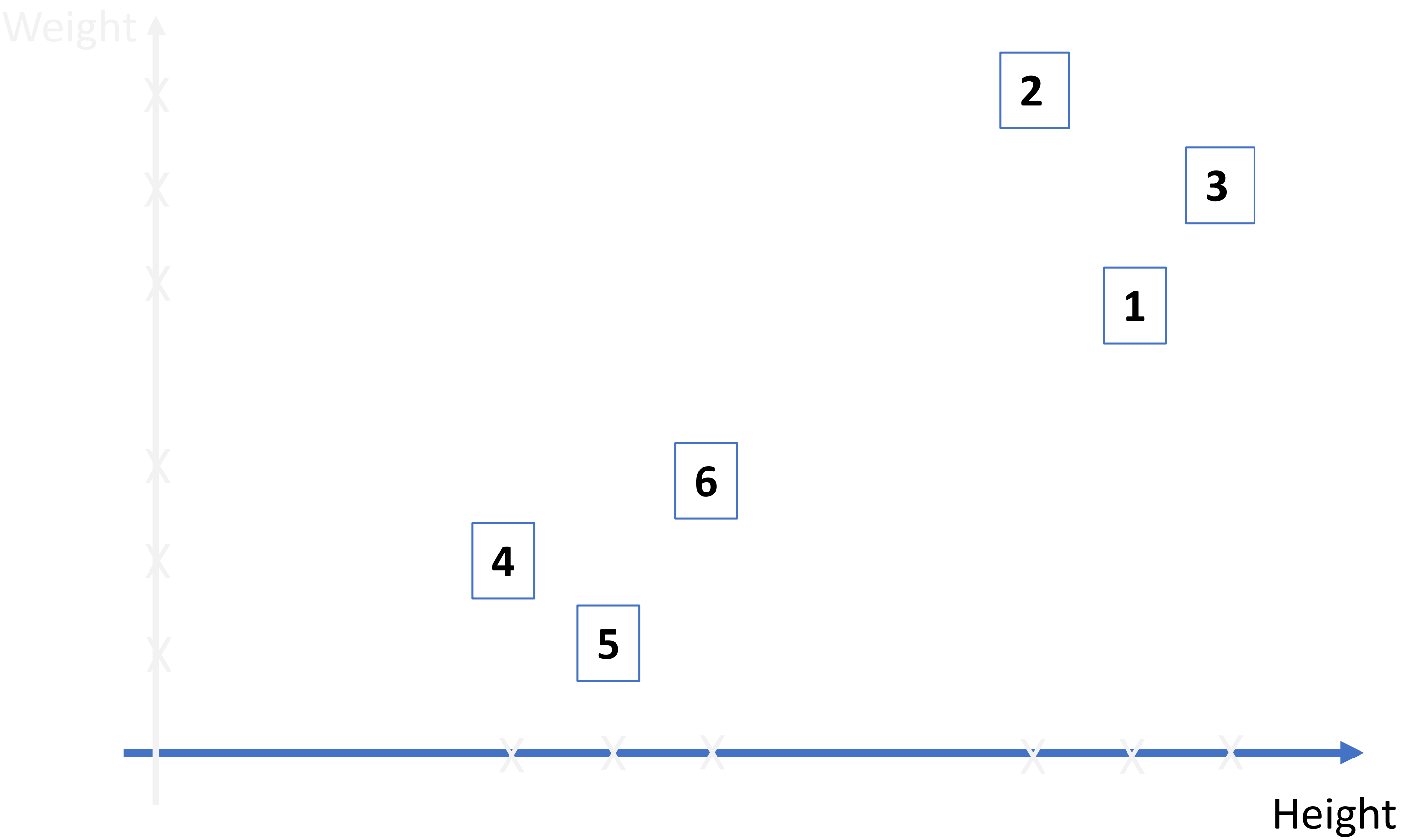
We can not visualize beyond 3 dimensions

	Height	Weight	#Wheel	CC
1	6	500	4	899
2	5.5	600	4	1000
3	6.5	550	4	800
4	3	200	2	99
5	3.5	150	2	125
6	4	250	2	100



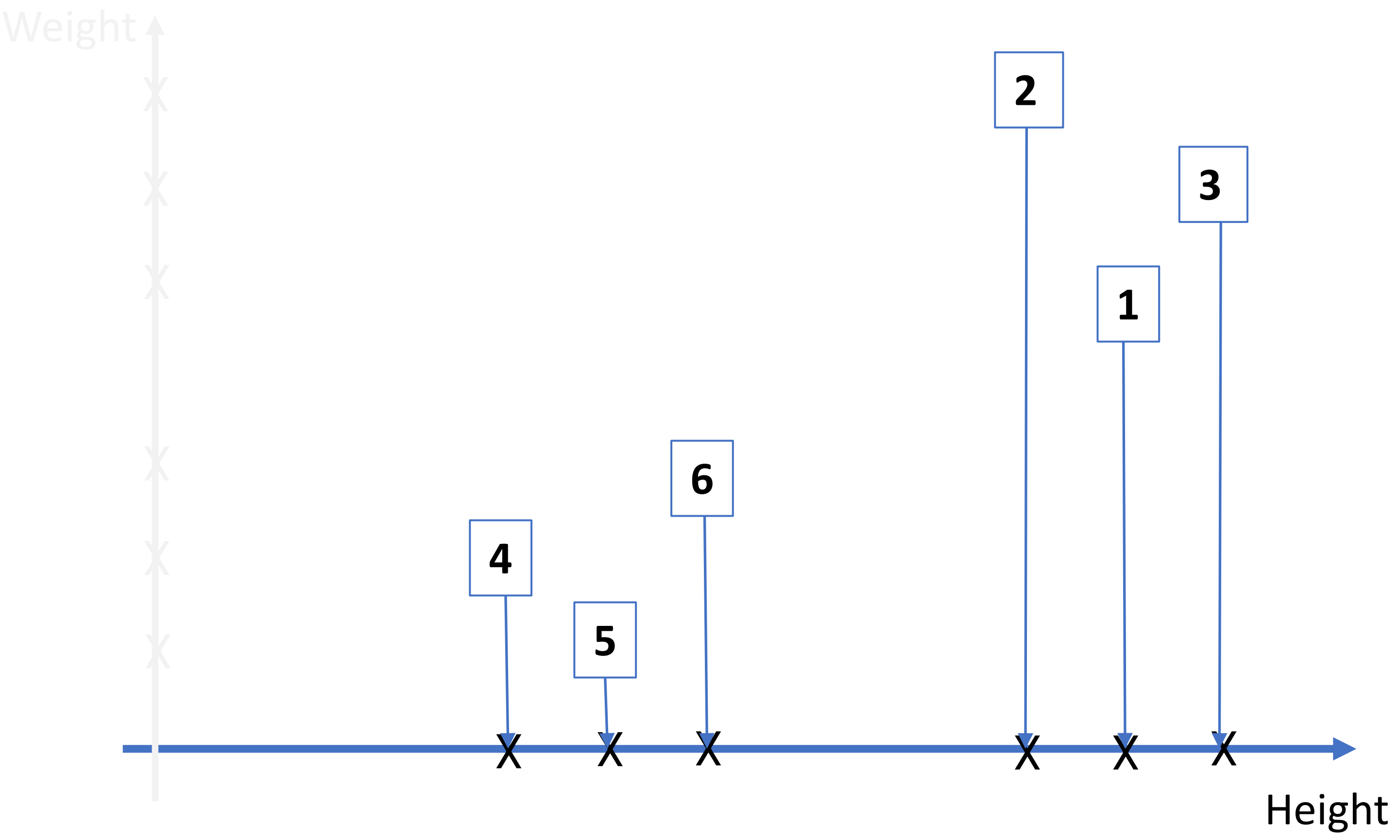
Representation of the data along X-Axis

	Height	Weight
1	6	500
2	5.5	600
3	6.5	550
4	3	200
5	3.5	150
6	4	250



Representation of the data along X-Axis

	Height	Weight
1	6	500
2	5.5	600
3	6.5	550
4	3	200
5	3.5	150
6	4	250



Let *D* be the dataset and

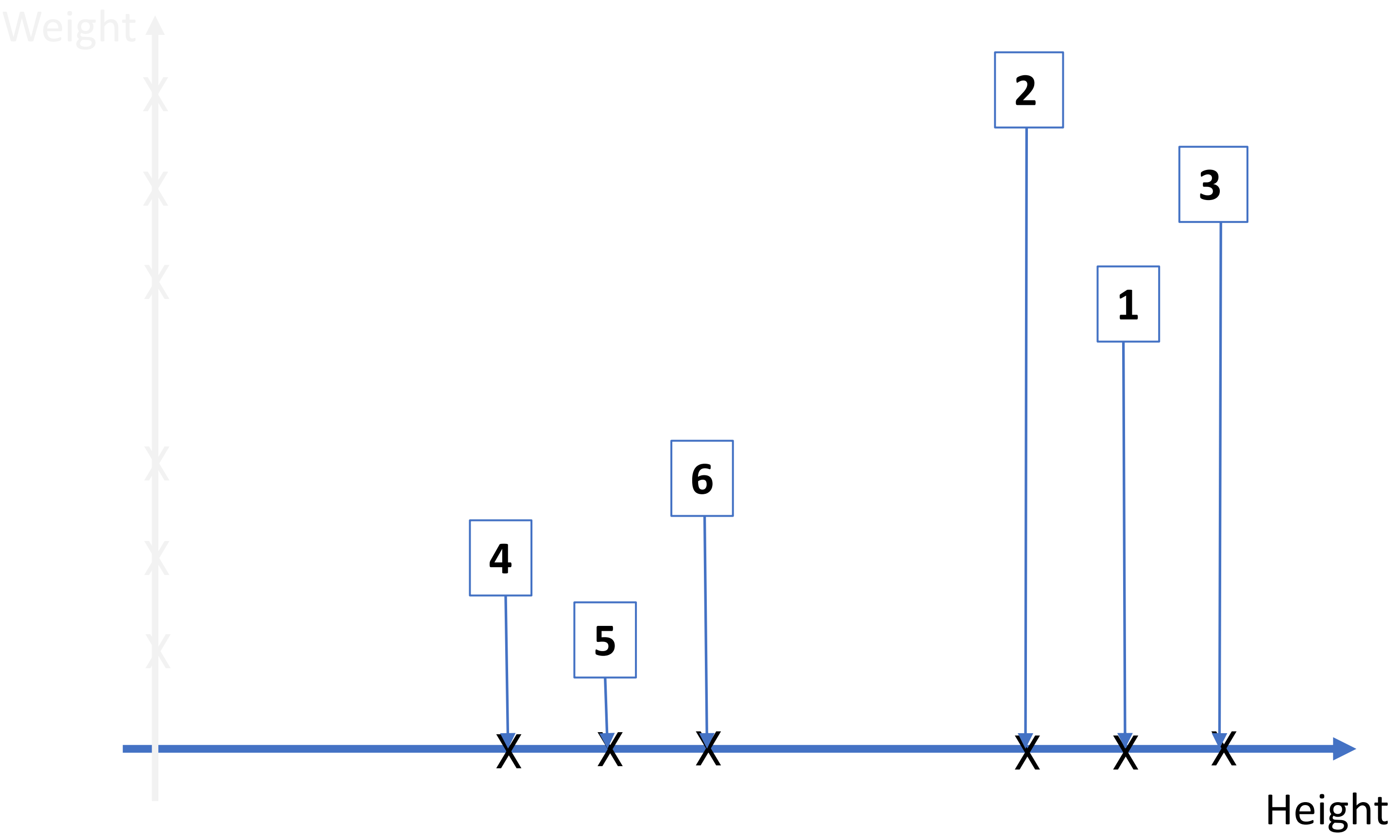
$\overline{x}^{\mathrm{T}}$ ={1 0} be the unit vector representing the x-axis

The projections of the data in *D* on x-axis are $D\overline{x}$

Representation of the data along X-Axis

	Height	Weight
1	6	500
2	5.5	600
3	6.5	550
4	3	200
5	3.5	150
6	4	250

$$\begin{bmatrix} 6 & 500 \end{bmatrix} \cdot \underset{\mathbf{x}}{\begin{bmatrix} 1 \\ 0 \end{bmatrix}} = \underset{D\overline{x}}{6}$$



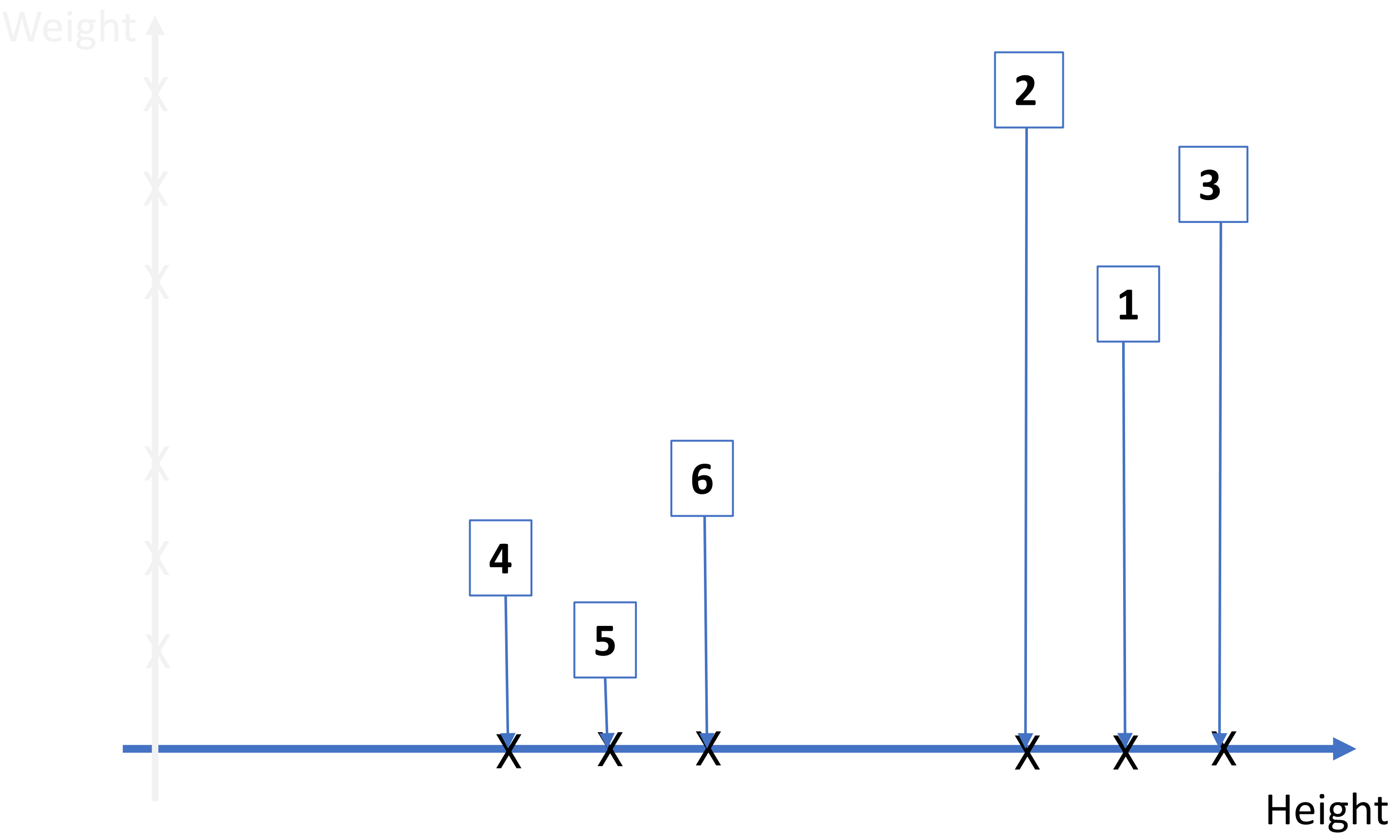
Let D be the dataset and

$\overline{x}^{\mathrm{T}}$ ={1 0} be the unit vector representing the x-axis

The projections of the data in D on x-axis are $D\overline{x}$

Representation of the data along X-Axis

	Height	Weight		x		$D\overline{x}$
1	6	500				6
2	5.5	600	•	1	=	5.5
3	6.5	550		0		
4	3	200				
5	3.5	150				
6	4	250				



Let D be the dataset and

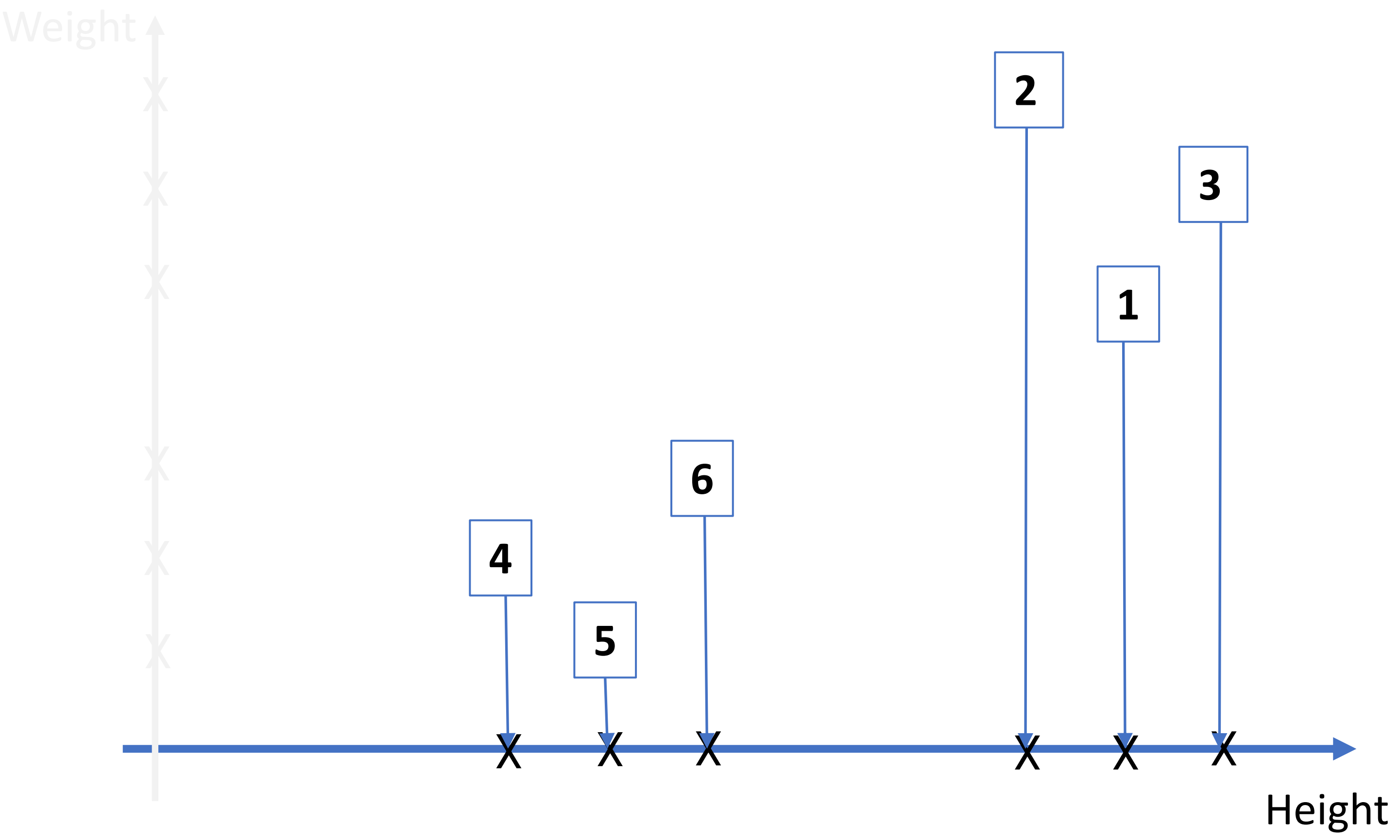
$\overline{x}^{\mathrm{T}}$ ={1 0} be the unit vector representing the x-axis

The projections of the data in D on x-axis are $D\overline{x}$

Representation of the data along X-Axis

	Height	Weight	$D\overline{x}$
1	6	500	6
2	5.5	600	5.5
3	6.5	550	6.5
4	3	200	
5	3.5	150	
6	4	250	

$$\begin{bmatrix} 6.5 & 550 \end{bmatrix} \cdot \mathbf{x} = \begin{bmatrix} 6.5 & 550 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 6.5$$



Let D be the dataset and

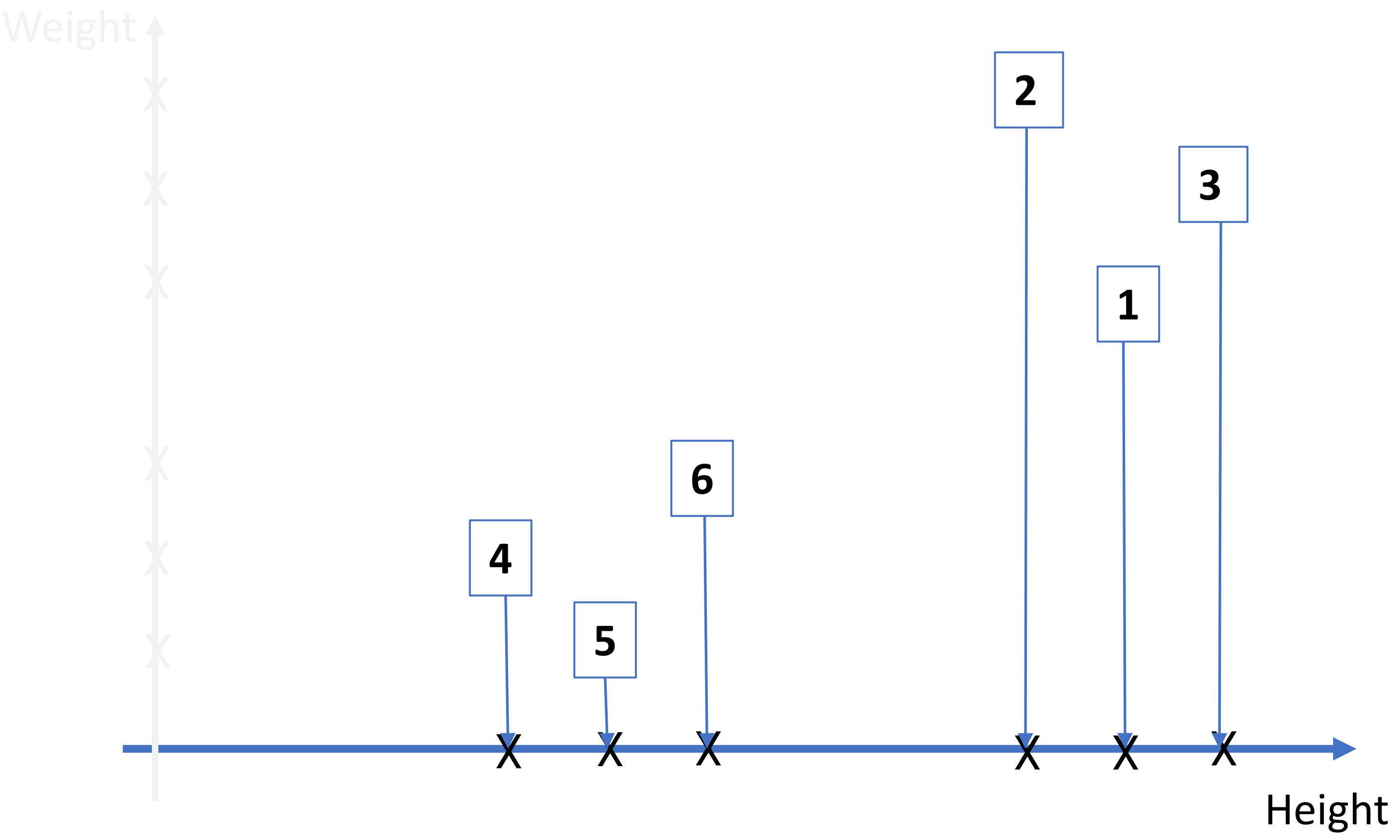
$\overline{x}^{\mathrm{T}}$ ={1 0} be the unit vector representing the x-axis

The projections of the data in D on x-axis are $D\overline{x}$

Representation of the data along X-Axis

	Height	Weight	$D\overline{x}$
1	**6**	**500**	6
2	**5.5**	**600**	5.5
3	**6.5**	**550**	6.5
4	**3**	**200**	3
5	**3.5**	**150**	
6	**4**	**250**	

$$\begin{bmatrix} 3 & 200 \end{bmatrix} \cdot \overset{\mathbf{x}}{\begin{bmatrix} 1 \\ 0 \end{bmatrix}} = 3$$



Let D be the dataset and

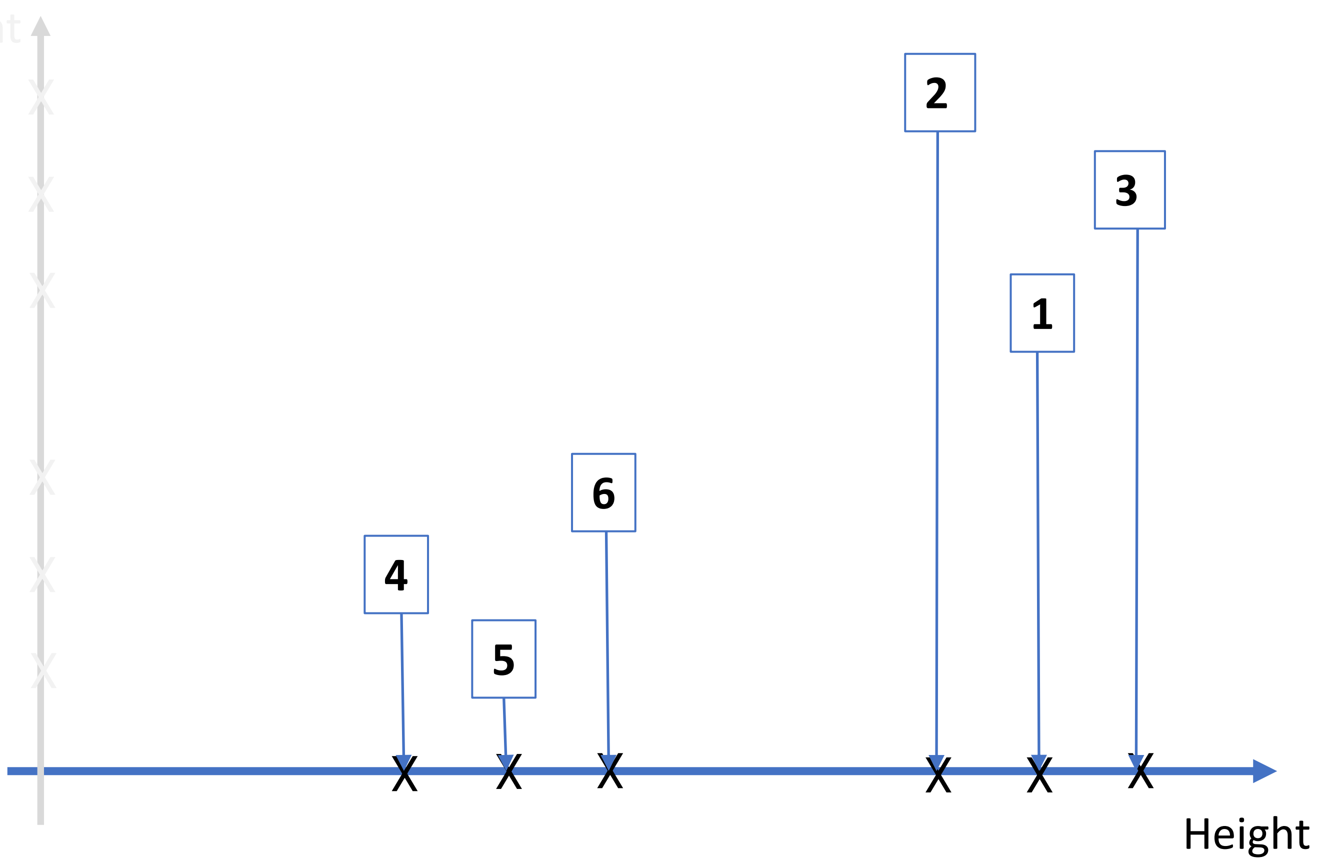
$\overline{x}^{\mathrm{T}}$ ={1 0} be the unit vector representing the x-axis

The projections of the data in D on x-axis are $D\overline{x}$

Representation of the data along X-Axis

	Height	Weight	$D\overline{x}$
1	6	500	6
2	5.5	600	5.5
3	6.5	550	6.5
4	3	200	3
5	3.5	150	3.5
6	4	250	

$$\begin{bmatrix} 3.5 & 150 \end{bmatrix} \cdot \underset{\mathbf{x}}{\begin{bmatrix} 1 \\ 0 \end{bmatrix}} = 3.5$$



Let D be the dataset and

$\overline{x}^{T}$ ={1 0} be the unit vector representing the x-axis

The projections of the data in D on x-axis are $D\overline{x}$

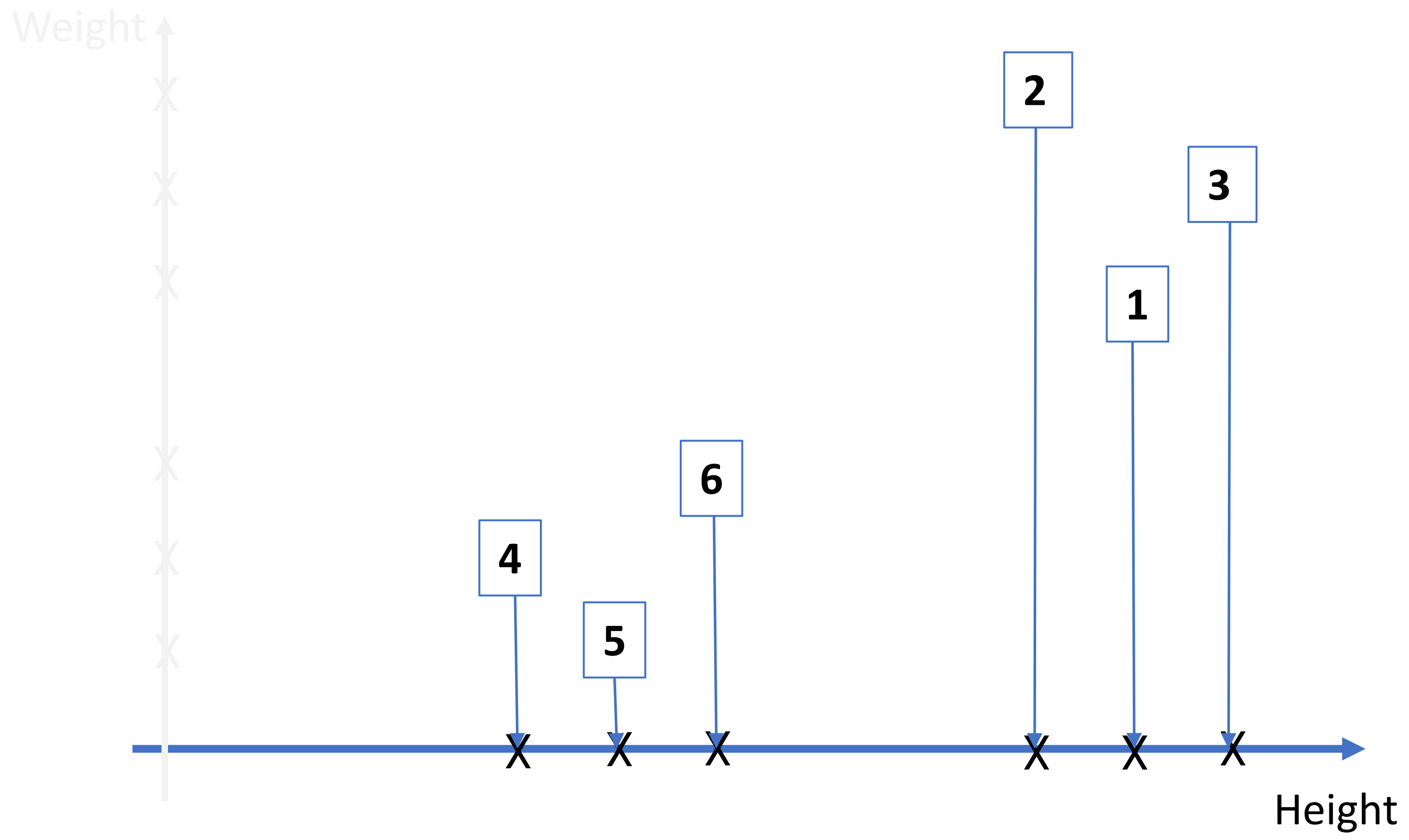
Representation of the data along X-Axis

	Height	Weight		$D\overline{x}$
1	6	500		6
2	5.5	600		5.5
3	6.5	550		6.5
4	3	200		3
5	3.5	150		3.5
6	4	250	$\cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ =	4

x

Let D be the dataset and

$\overline{x}^{\mathrm{T}}$ ={1 0} be the unit vector representing the x-axis

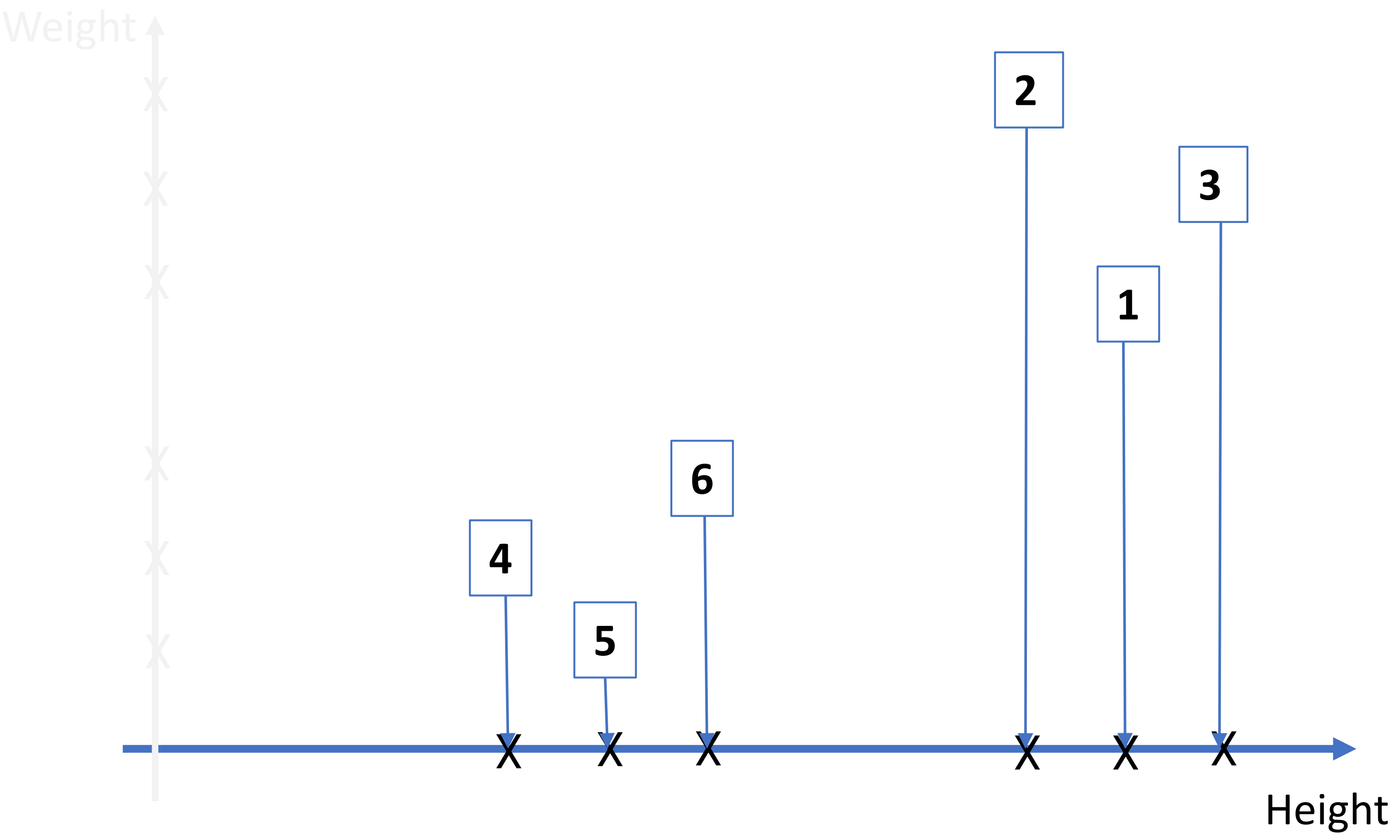


The projections of the data in D on x-axis are $D\overline{x}$

Representation of the data along X-Axis

	Height	**Weight**
1	6	500
2	5.5	600
3	6.5	550
4	3	200
5	3.5	150
6	4	250

$D\overline{x}$
6
5.5
6.5
3
3.5
4



Let D be the dataset and

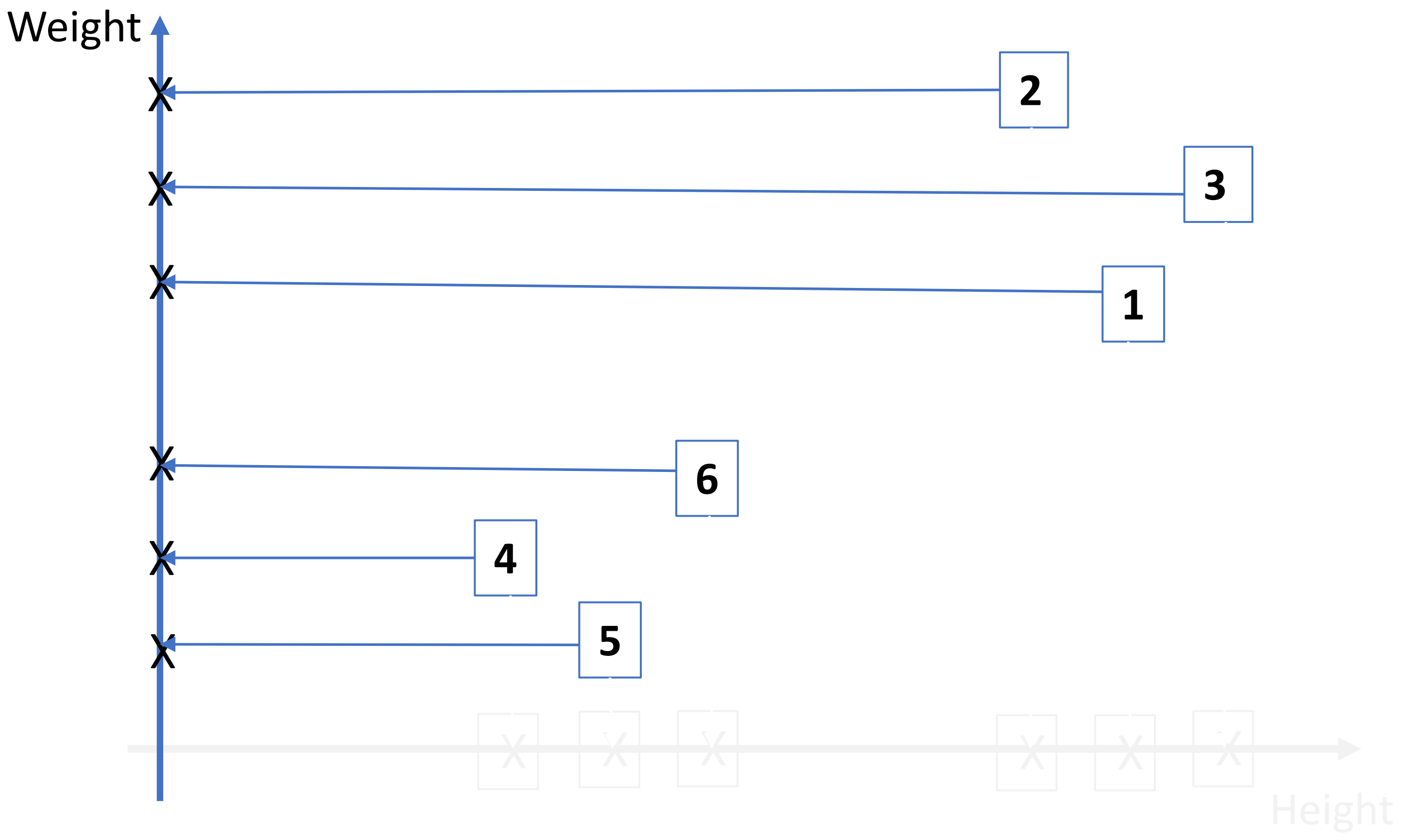
$\overline{x}^{\mathrm{T}}$ ={1 0} be the unit vector representing the x-axis

The projections of the data in D on x-axis are $D\overline{x}$

Representation of the data along Y-Axis

	Height	Weight
1	6	500
2	5.5	600
3	6.5	550
4	3	200
5	3.5	150
6	4	250

$D\overline{y}$
6
5.5
6.5
3
3.5
4



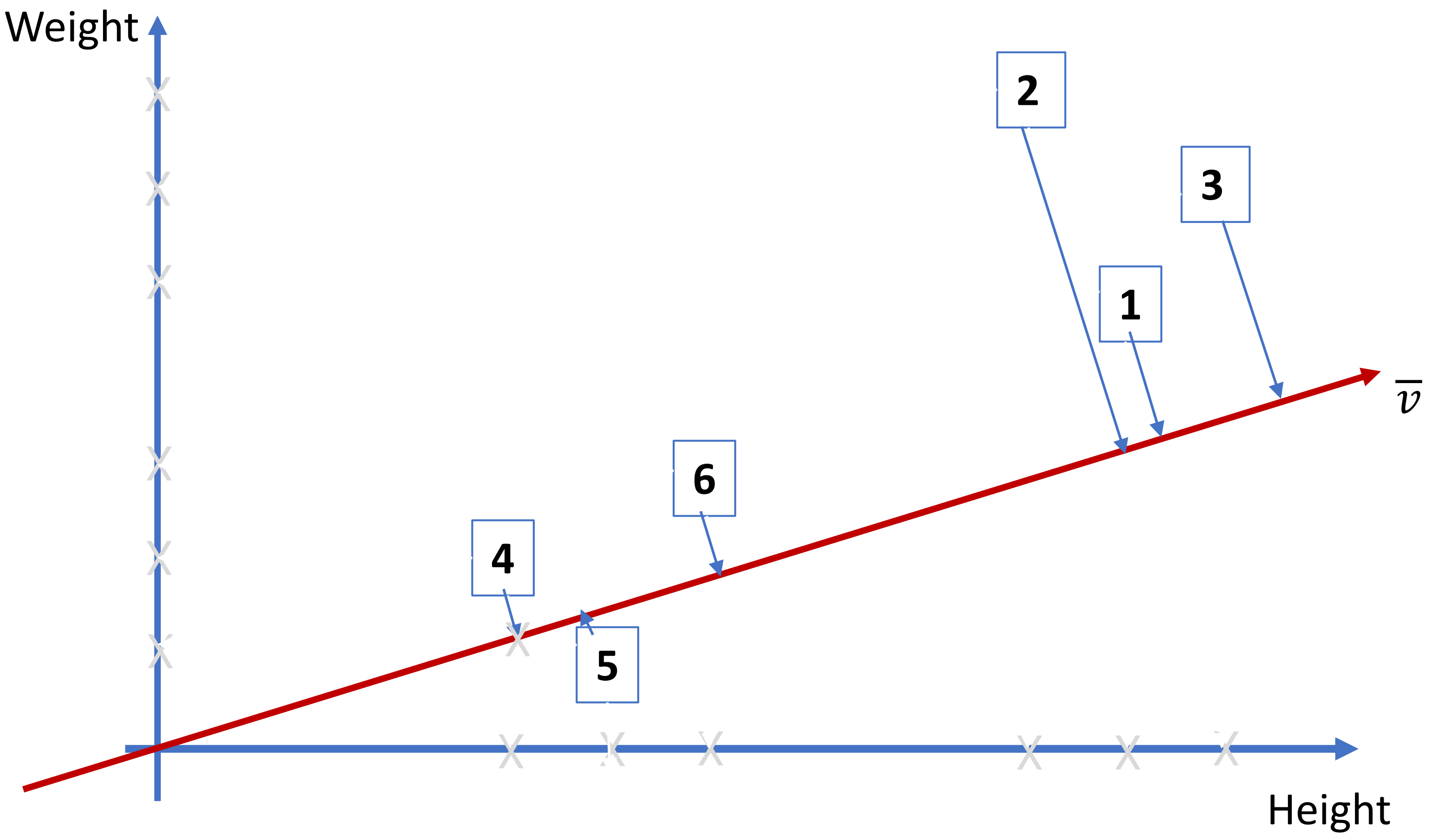
Let D be the dataset and

$\overline{y}^{\mathrm{T}}$ ={0 1} be the unit vector representing the y-axis

The projections of the data in D on y-axis are $D\overline{y}$

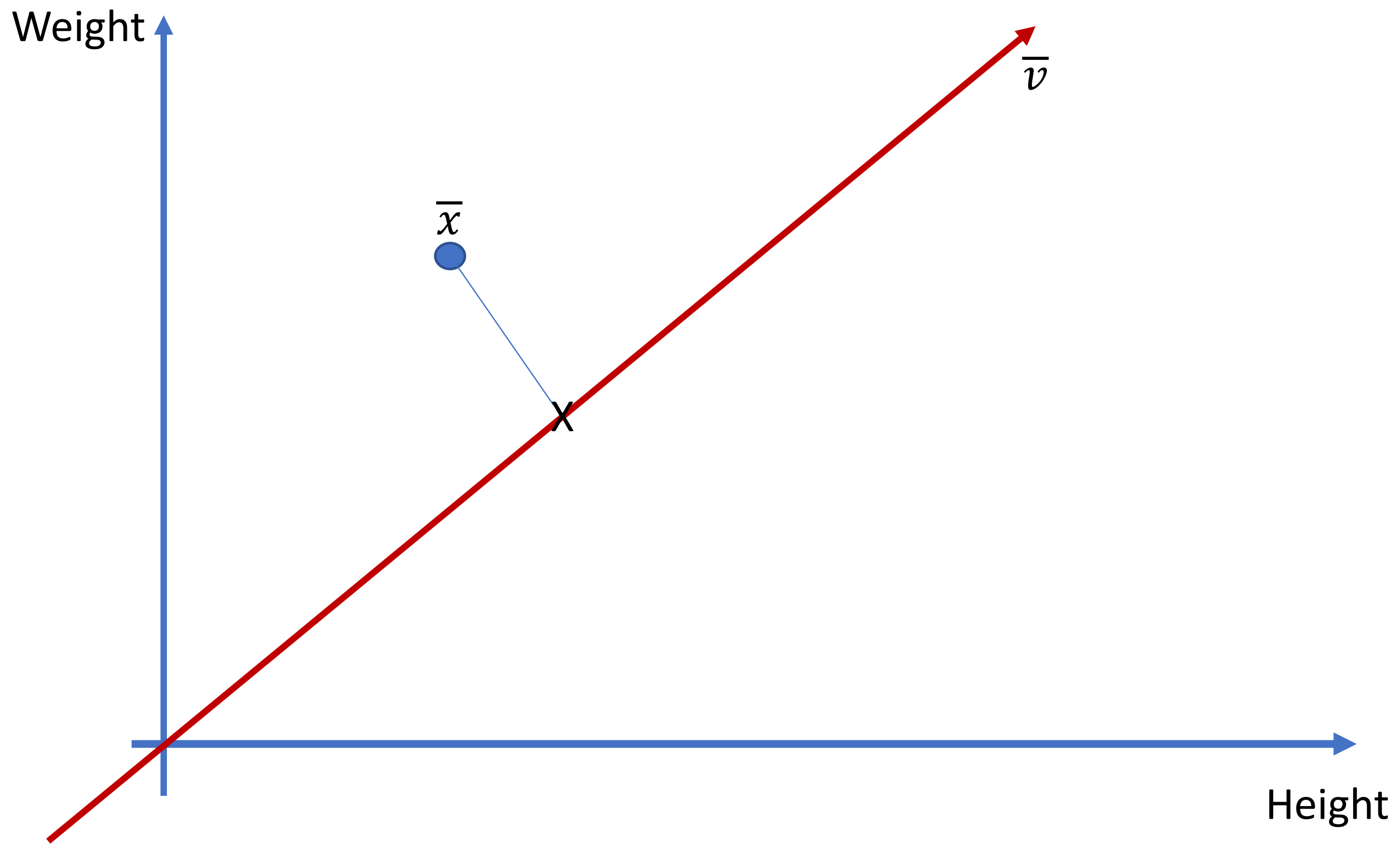
Representation of the data along a random vector

	Height	Weight
1	6	500
2	5.5	600
3	6.5	550
4	3	200
5	3.5	150
6	4	250



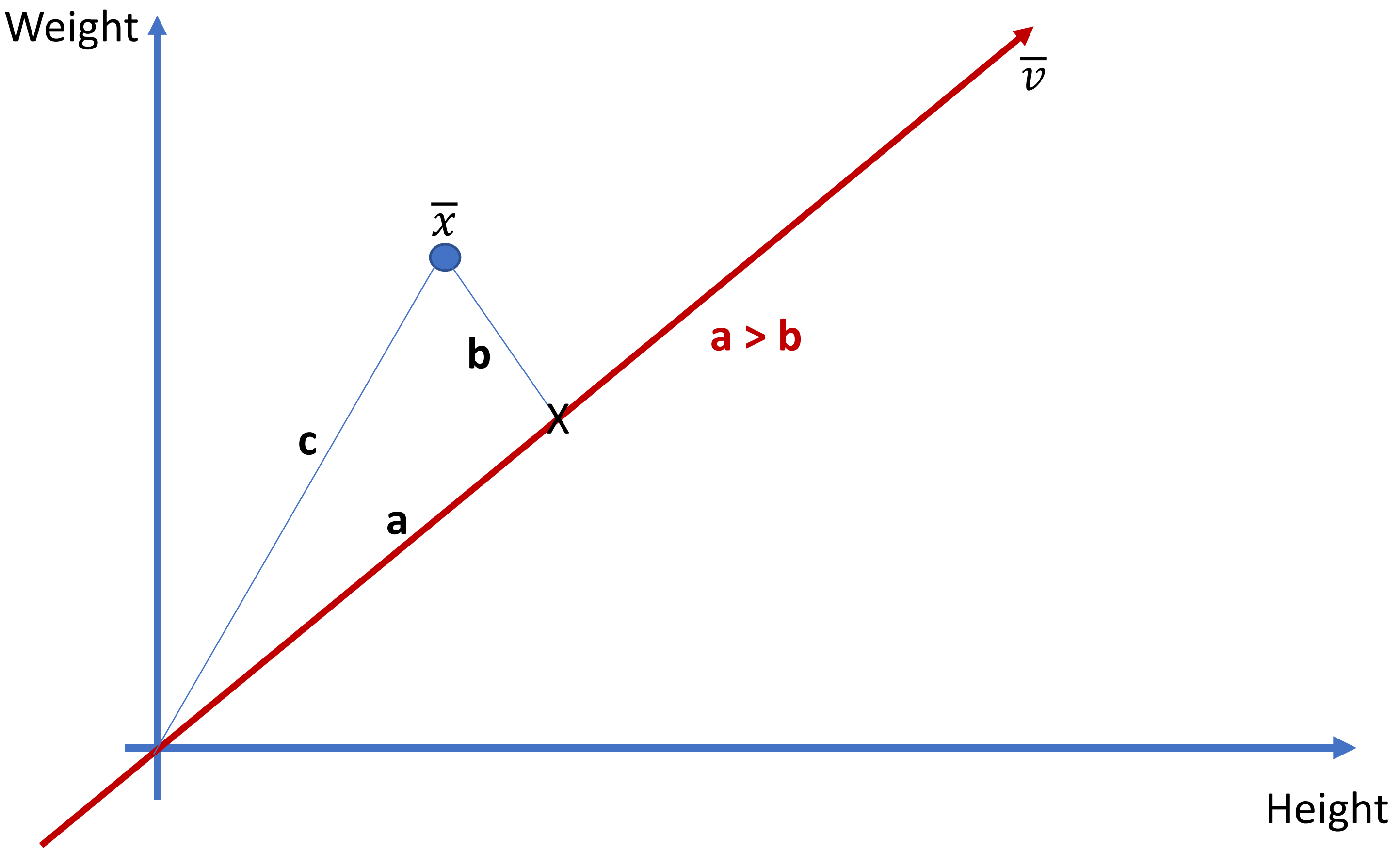
The projections of the data in *D* on
A random direction $\overline{v}$ are $D\overline{v}$

Projection of a point on a random vector



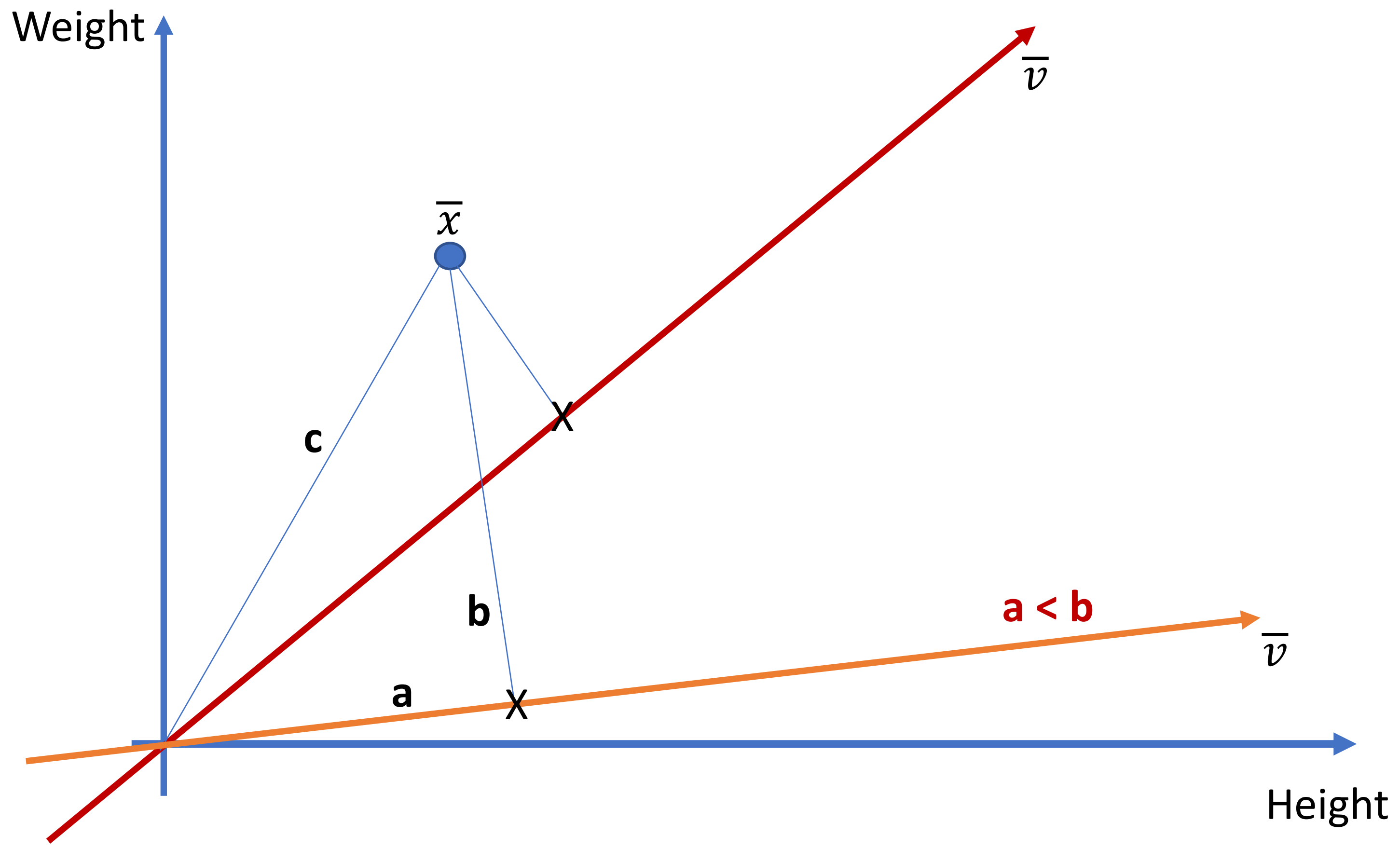
Projection of a point on a random vector

$c^2 = a^2 + b^2$



Projection of a point on a random vector

$c^2 = a^2 + b^2$

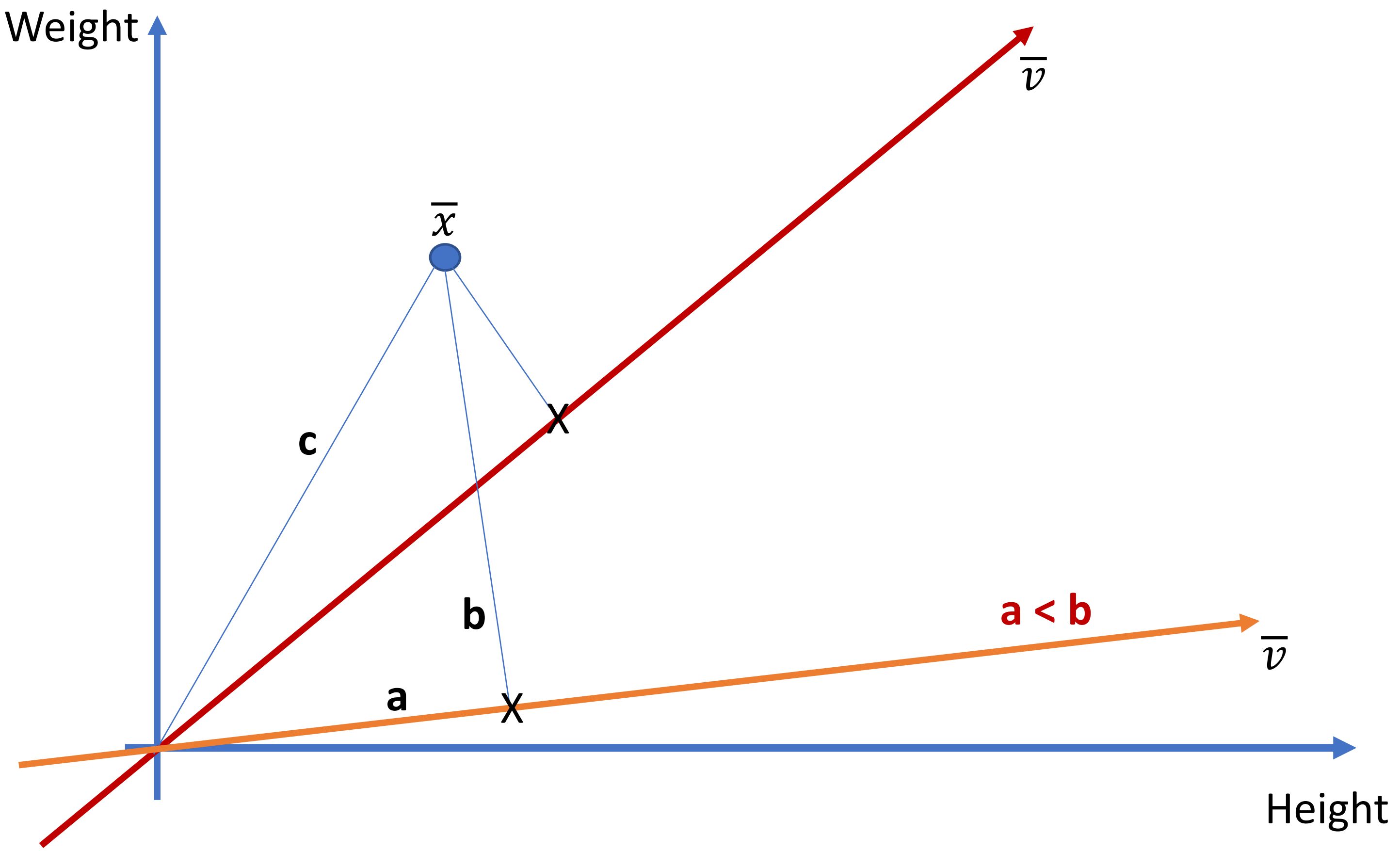


Projection of a point on a random vector

$c^2 = a^2 + b^2$

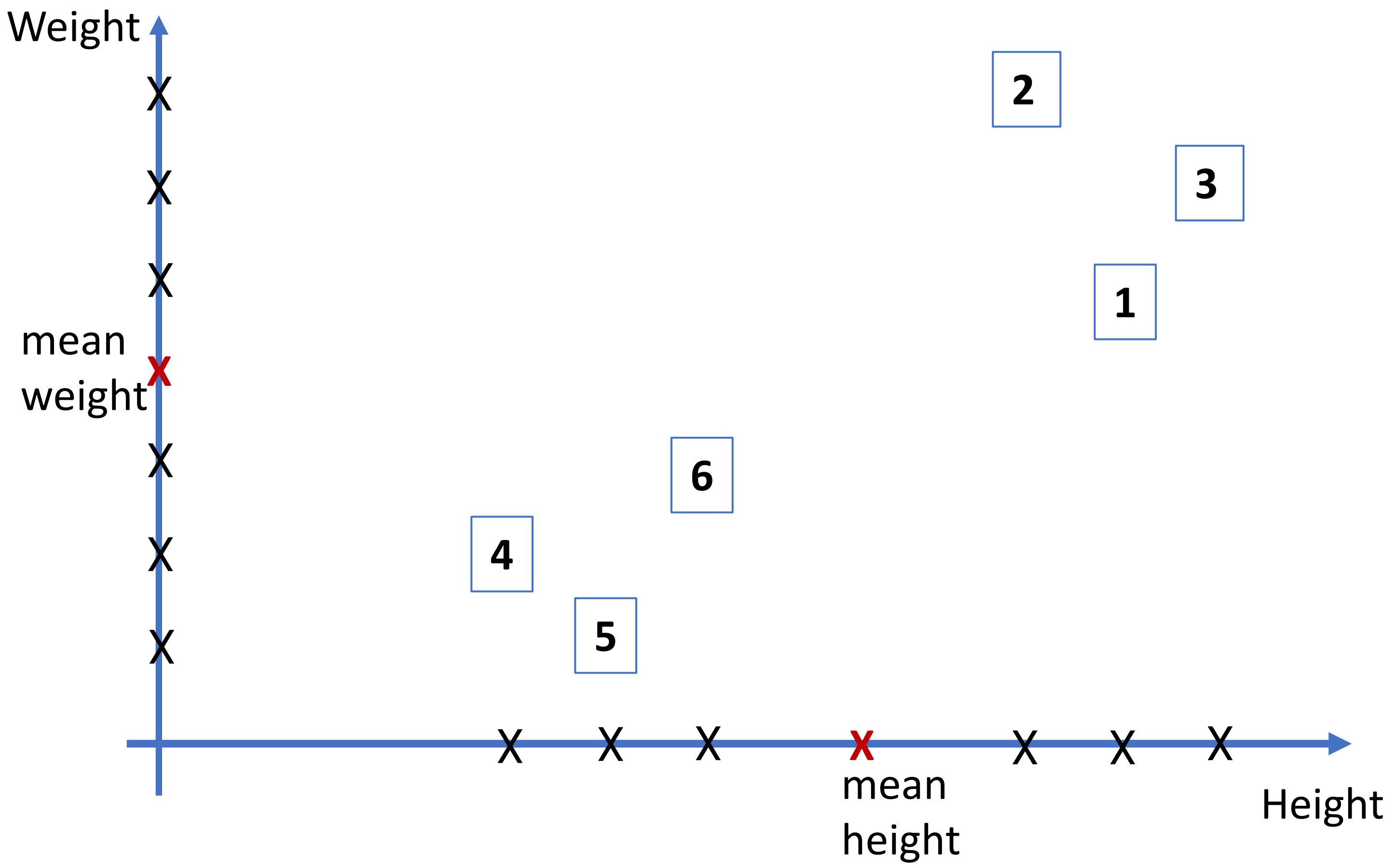
As *b* increases,
**   *a* decreases**

and vice versa



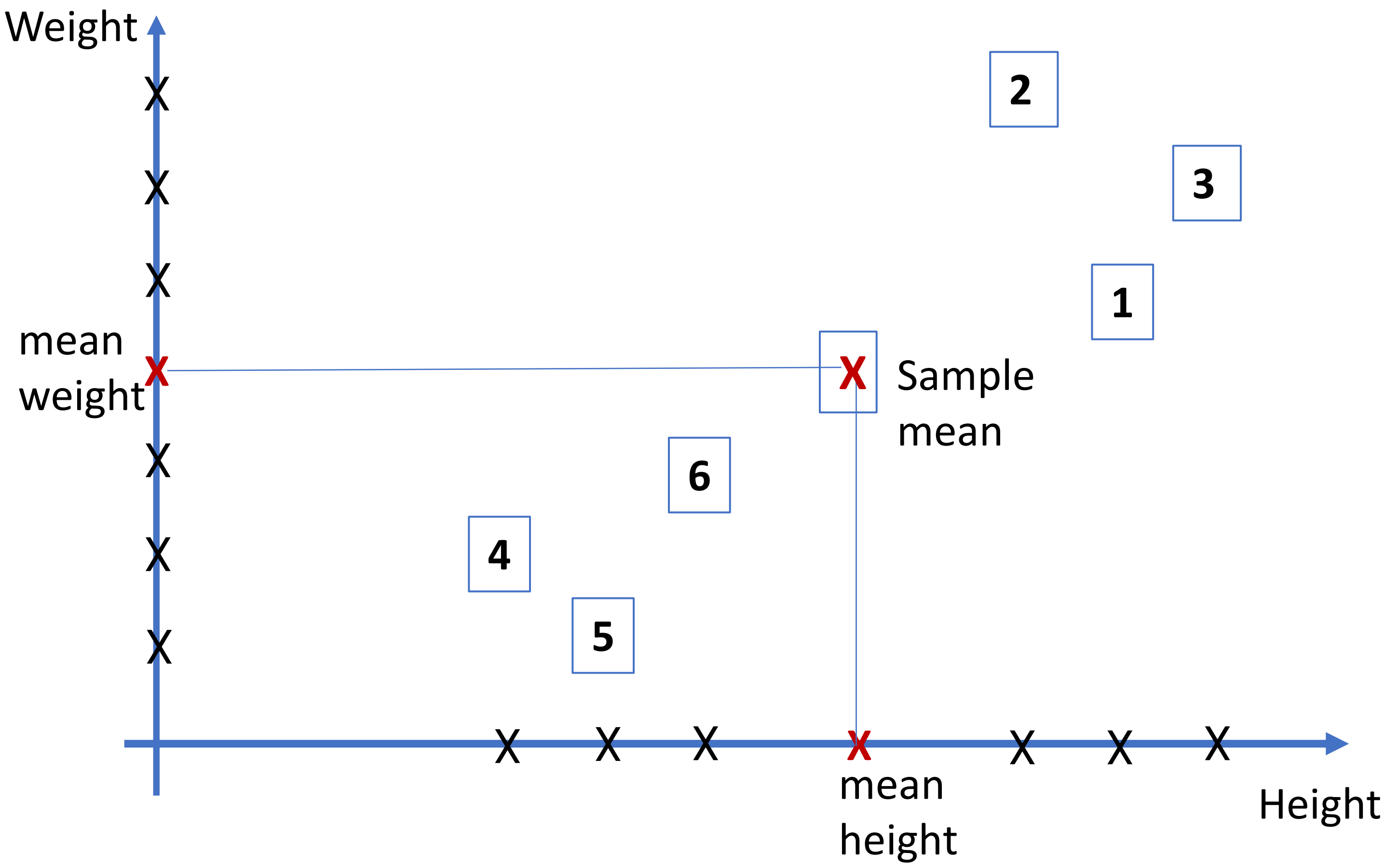
Sample Mean

	Height	Weight
1	6	500
2	5.5	600
3	6.5	550
4	3	200
5	3.5	150
6	4	250



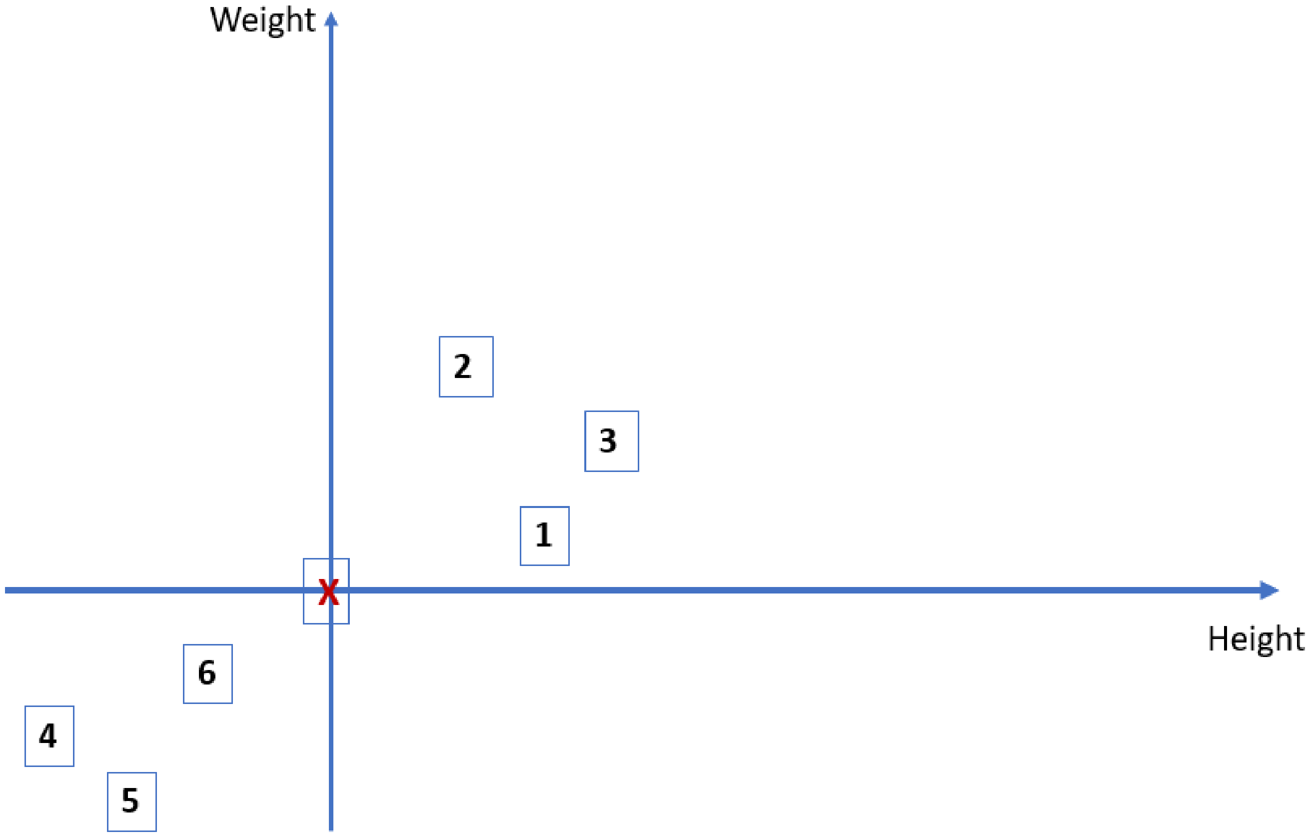
Sample Mean

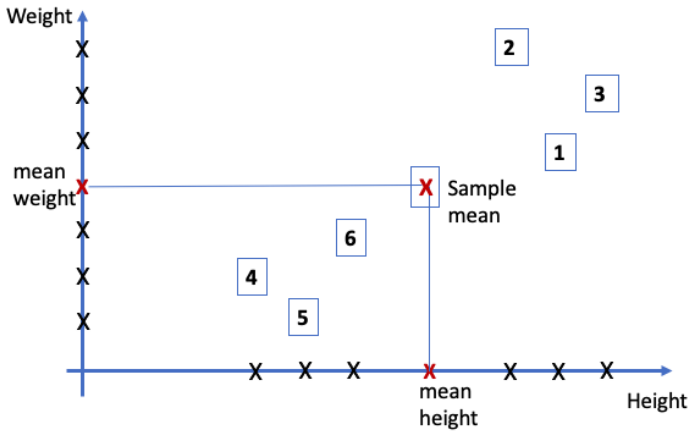
	Height	Weight
1	6	500
2	5.5	600
3	6.5	550
4	3	200
5	3.5	150
6	4	250



Shifting Mean

	Height	Weight
1	6	500
2	5.5	600
3	6.5	550
4	3	200
5	3.5	150
6	4	250

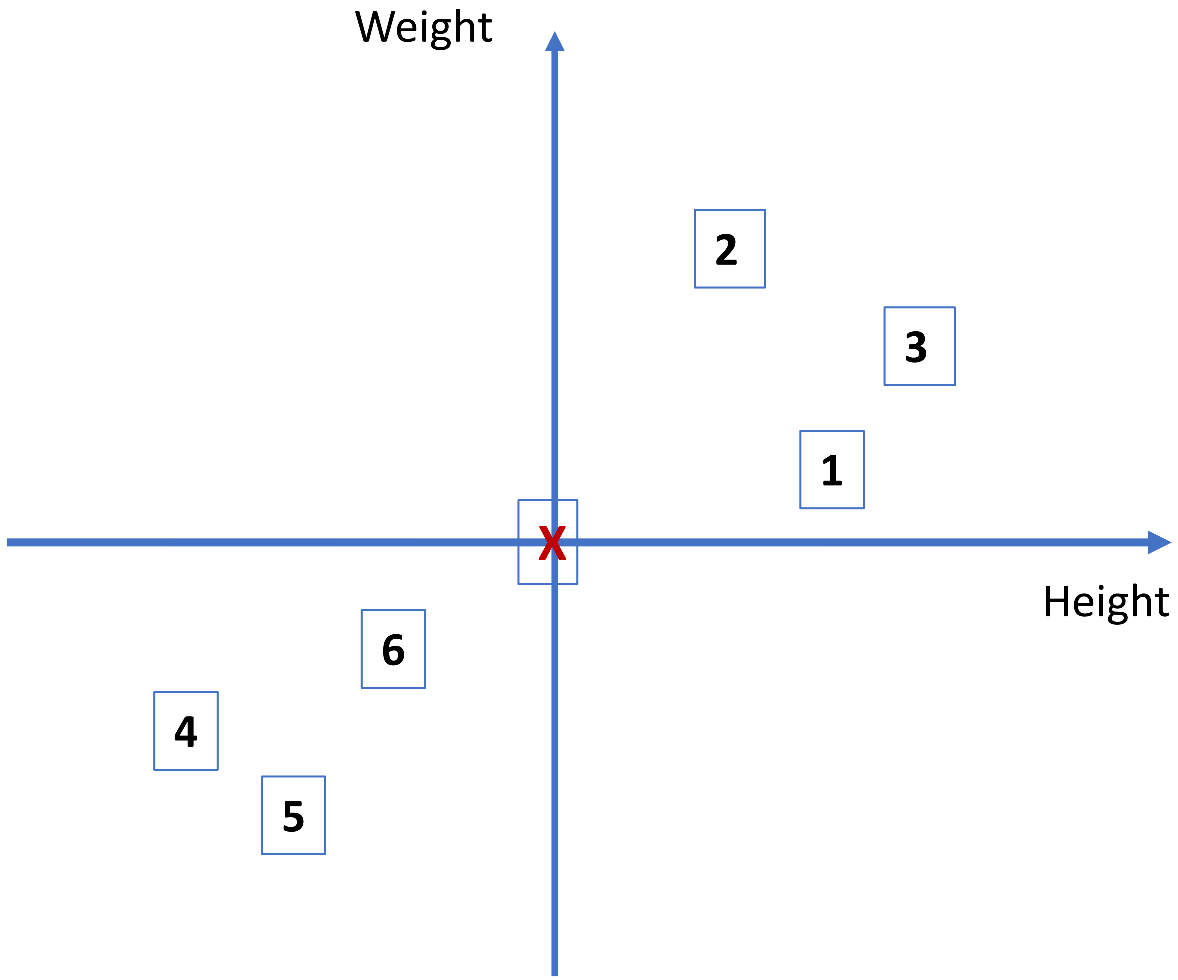




Best Fit?

Mean adjusted

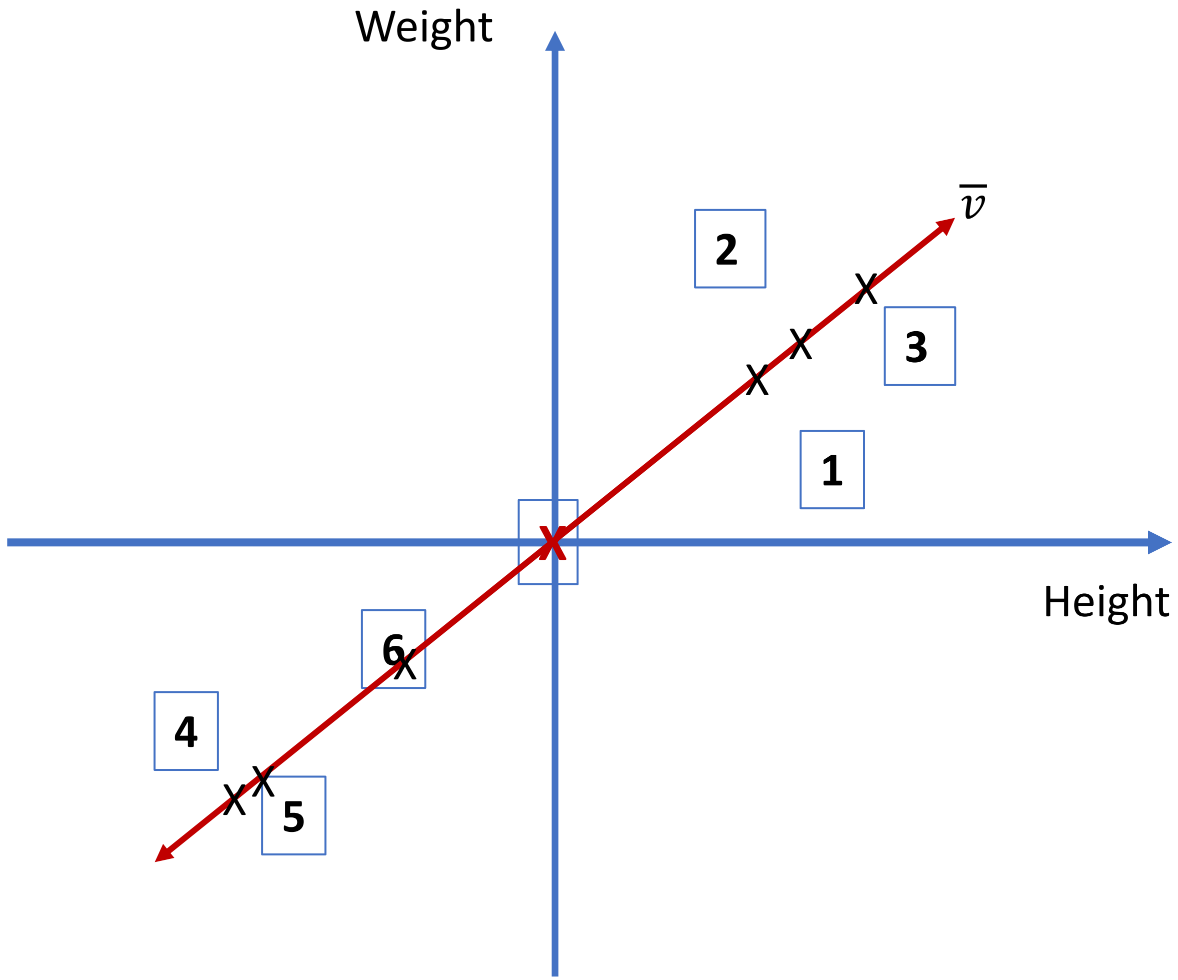
h - m	w - m
1.25	125
0.75	225
1.75	175
-1.75	-175
-1.25	-225
-0.75	-125



Best Fit

Mean adjusted

h - m	w - m
1.25	125
0.75	225
1.75	175
-1.75	-175
-1.25	-225
-0.75	-125



Best Fit

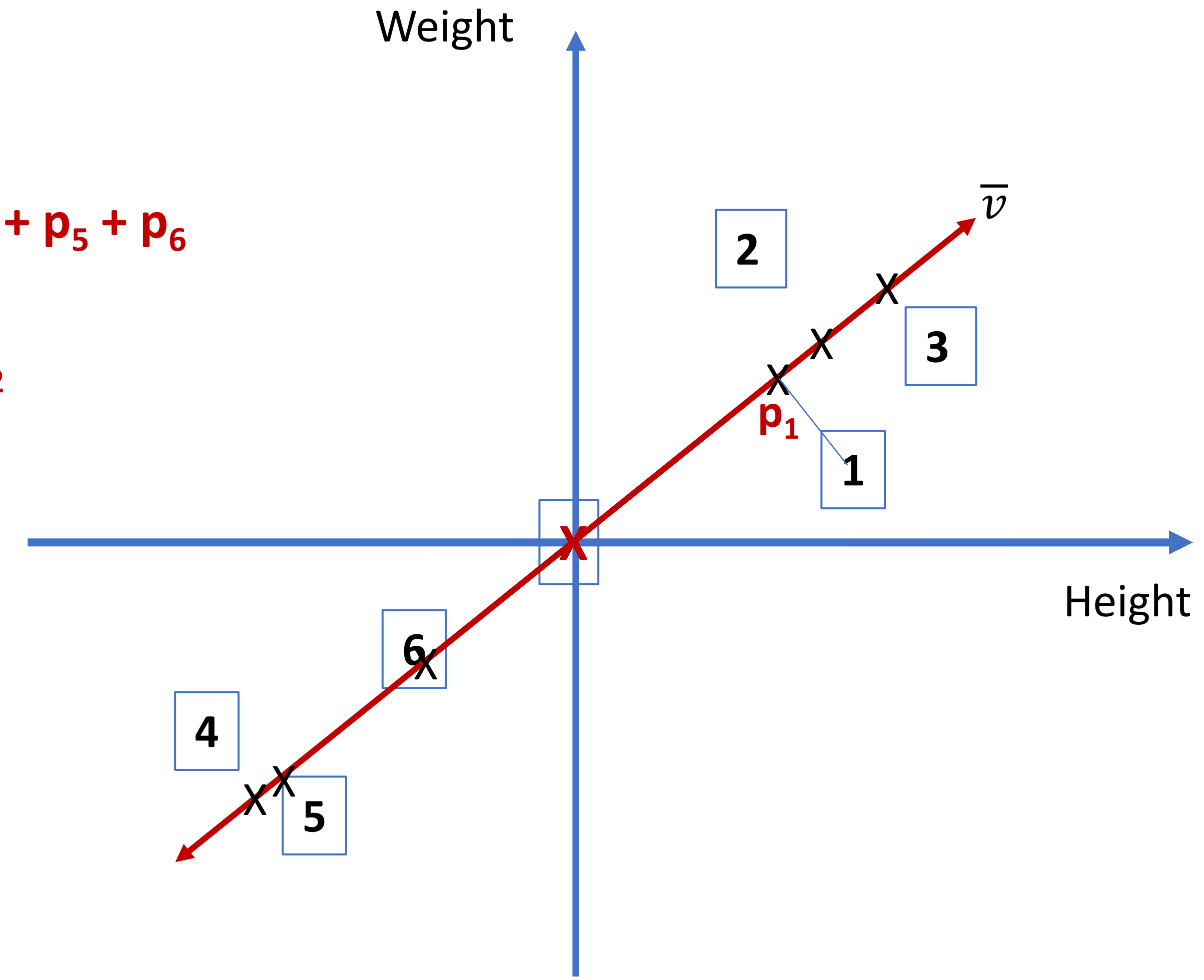
Mean adjusted

h - m	w - m
1.25	125
0.75	225
1.75	175
-1.75	-175
-1.25	-225
-0.75	-125

Sum of Error = $p_1 + p_2 + p_3 + p_4 + p_5 + p_6$

Sum of Square Error
$= p_1^2 + p_2^2 + p_3^2 + p_4^2 + p_5^2 + p_6^2$

Vector with the minimum sum of square error is the Best fit vector.



Best Fit

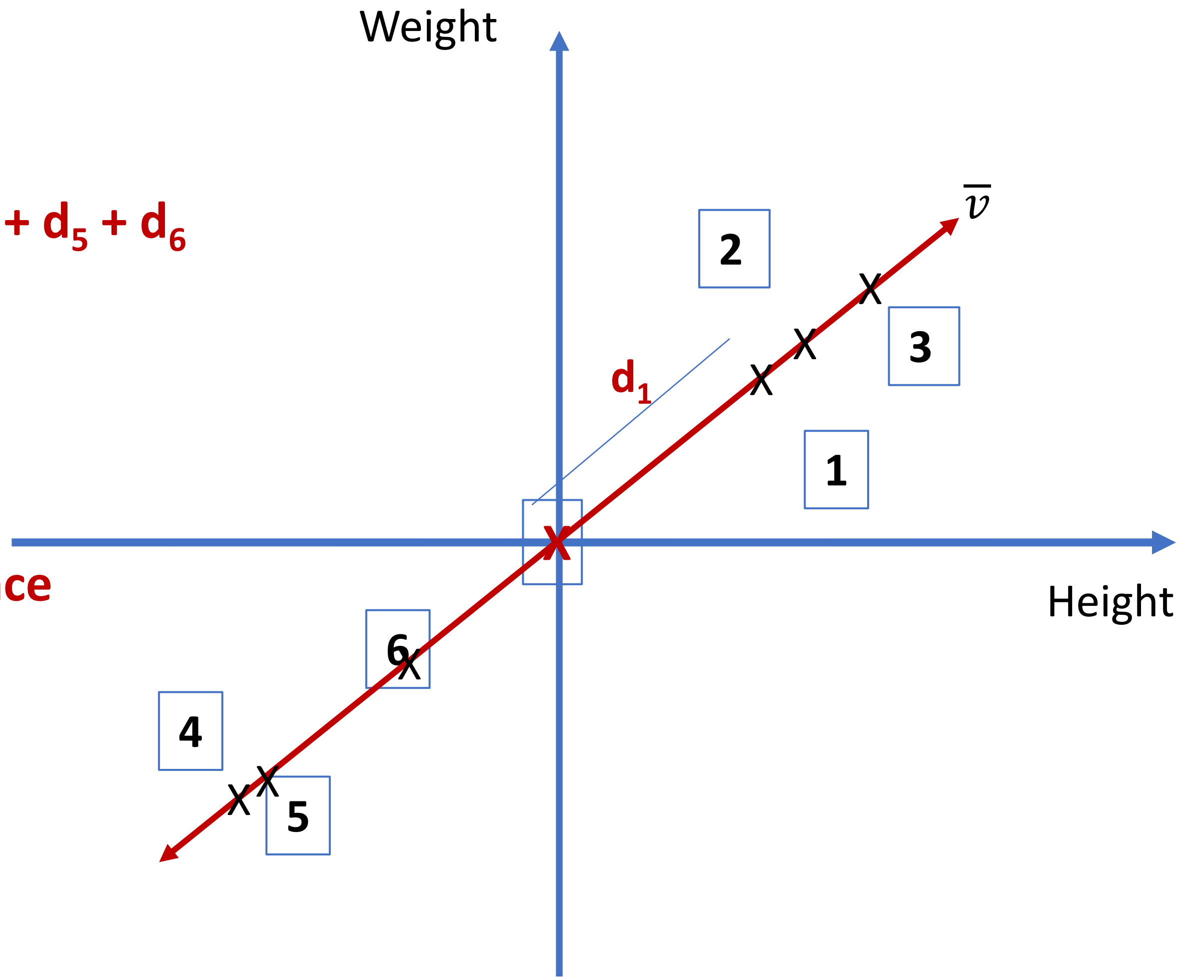
Mean adjusted

h - m	w - m
1.25	125
0.75	225
1.75	175
-1.75	-175
-1.25	-225
-0.75	-125

Sum distance = $d_1 + d_2 + d_3 + d_4 + d_5 + d_6$

Sum of square distance
= $d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2$

Vector with the
maximum sum of square distance
is the Best fit vector.



Principal Components

Mean adjusted

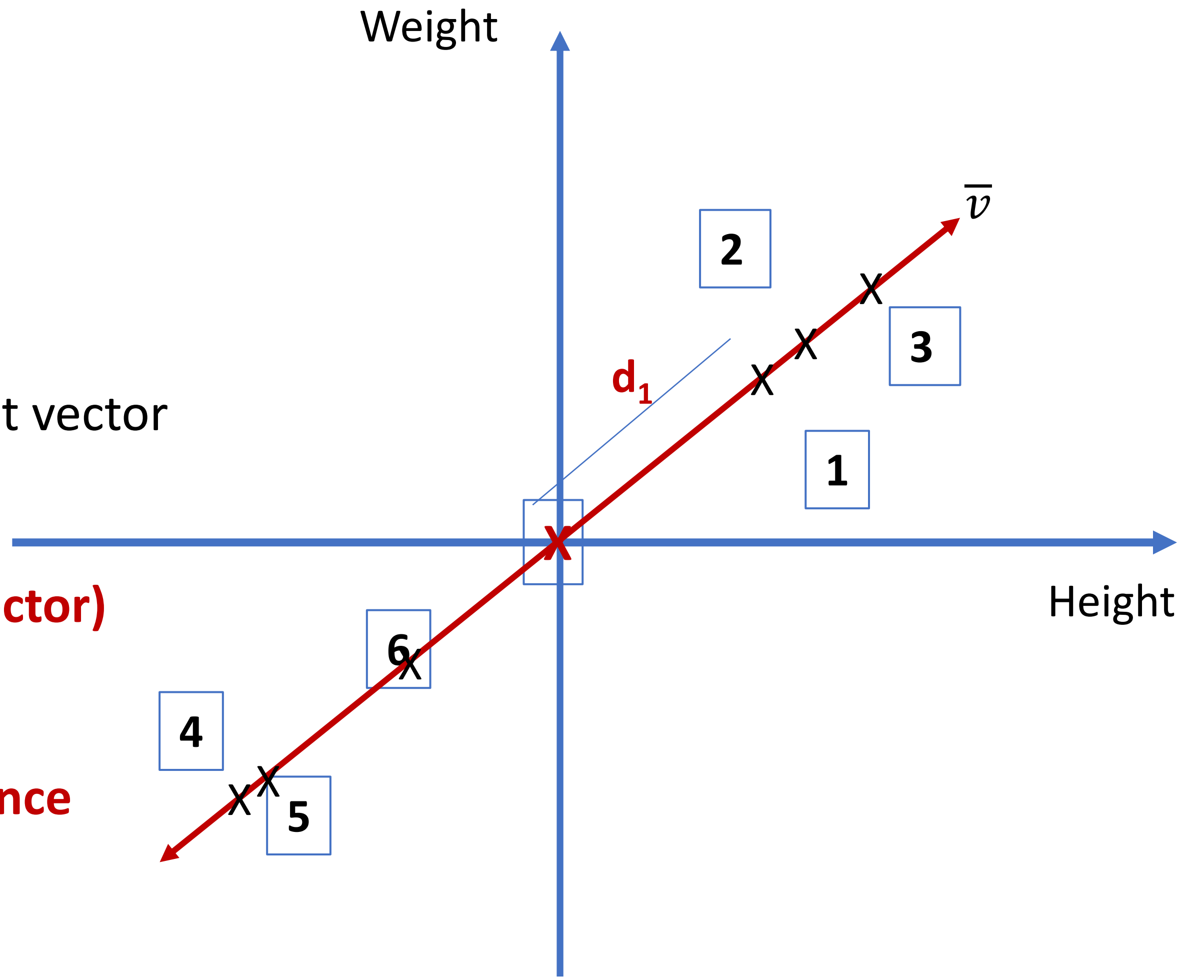
h - m	w - m
1.25	125
0.75	225
1.75	175
-1.75	-175
-1.25	-225
-0.75	-125

Sum of square distance
$= d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2$

Let us say , $\overline{v}$ is the best fitting unit vector

$\overline{v}$ is the called
1st Principal Component (eigenvector)

The mean of sum of square distance (variance)
is called Eigenvalue
of the 1st Principal Component



Principal Components

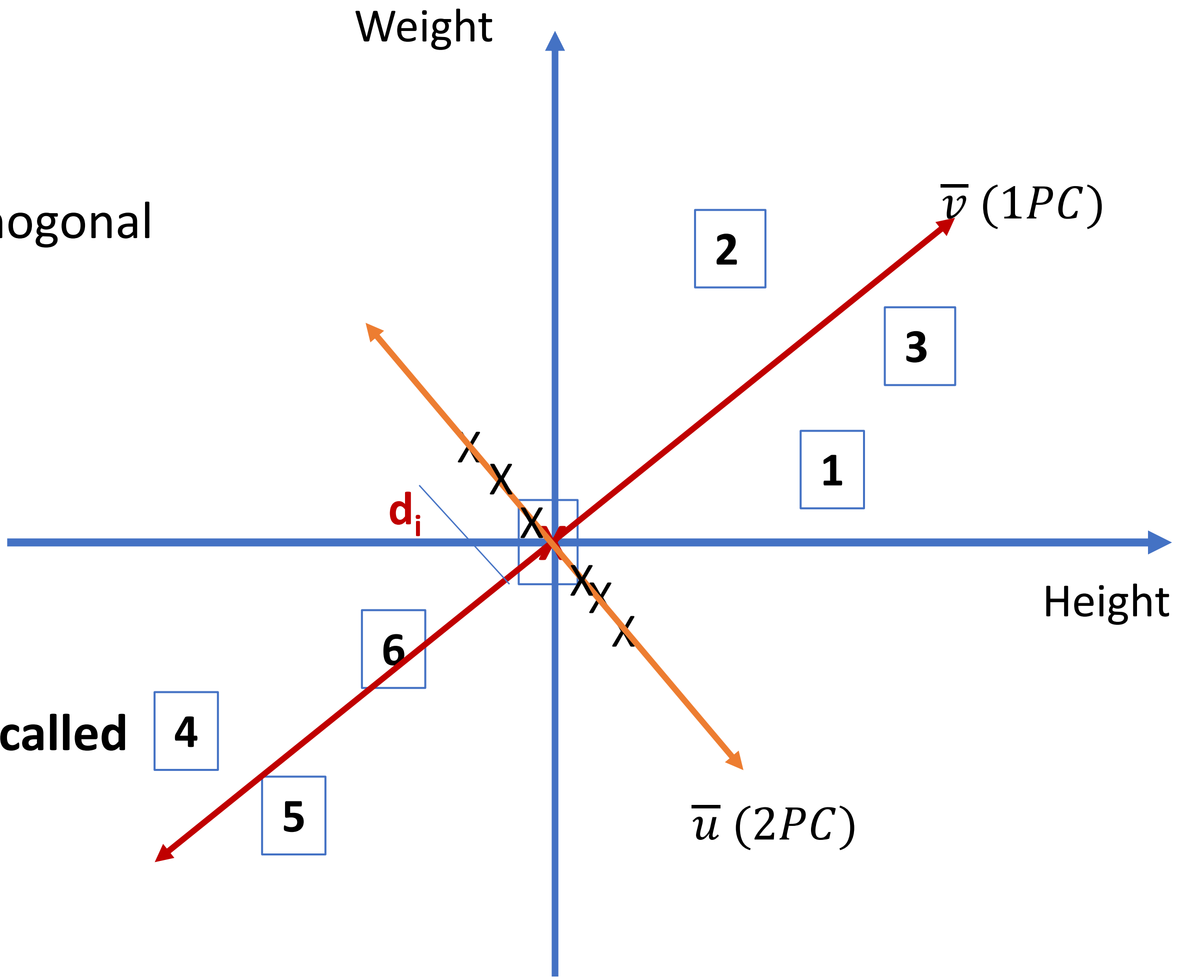
Mean adjusted

h - m	w - m
1.25	125
0.75	225
1.75	175
-1.75	-175
-1.25	-225
-0.75	-125

Let $\overline{u}$ be a vector the vector orthogonal
to
vector $\overline{v}$. Then, $\overline{u}$ is called
2nd Principal Component

Sum of square distance
= $d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2$

The mean of square distance is called Eigenvalue of the 2nd Principal Component



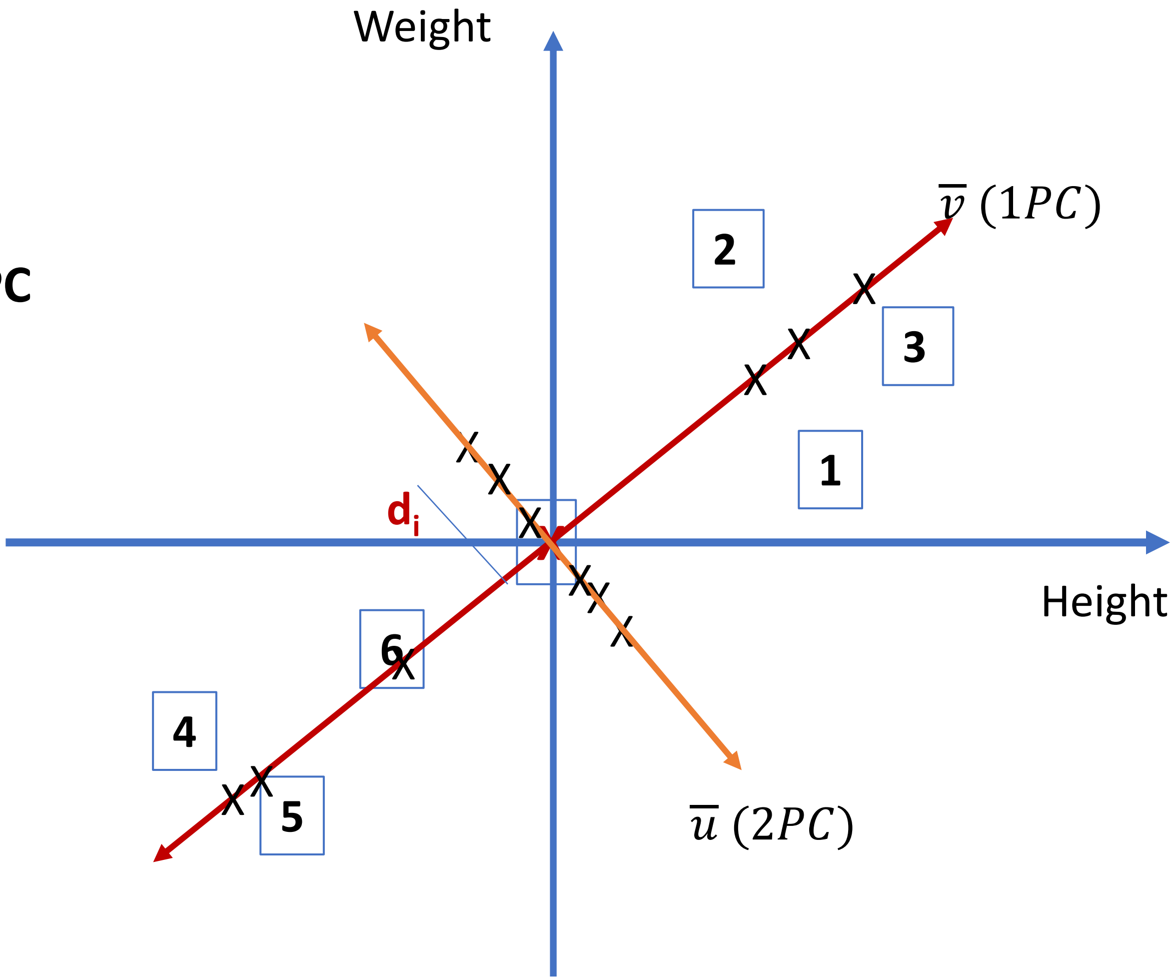
Principal Components

Mean adjusted

h - m	w - m
1.25	125
0.75	225
1.75	175
-1.75	-175
-1.25	-225
-0.75	-125

Large the variance,
more is the importance of the PC

1PC is more important than the 2PC

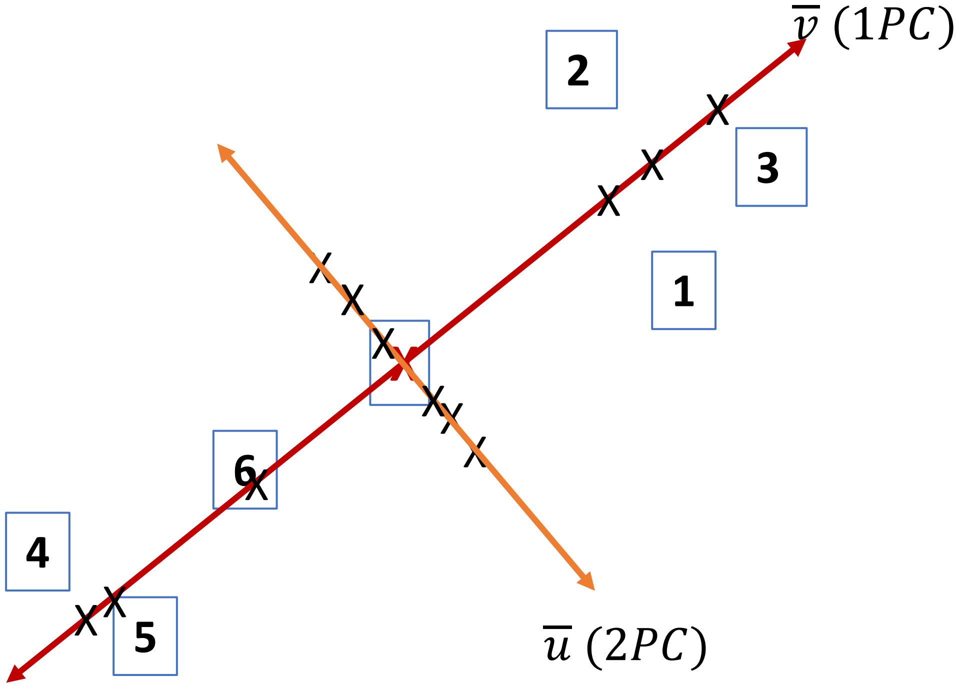


Data points on Principal Components

Mean adjusted

h - m	w - m
1.25	125
0.75	225
1.75	175
-1.75	-175
-1.25	-225
-0.75	-125

1PC	2PC
x11	x21
x12	x22
x13	x23
x14	x24
x15	x25
x16	x26



Data points on Principal Components

Dataset

y_1	y_2	y_3		y_d
y_{11}	y_{21}	y_{31}	….	y_{d1}
y_{12}	y_{22}	y_{32}		y_{d2}
y_{13}	y_{23}	y_{33}		y_{d3}
y_{14}	y_{24}	y_{34}		y_{d4}
.	.	.		.
y_{1n}	y_{2n}	y_{3n}		y_{dn}

Mean adjusted

$y_1 - m_1$	$y_2 - m_2$	$y_3 - m_3$		$y_d - m_d$
ma_{11}	ma_{21}	ma_{31}	….	ma_{d1}
ma_{12}	ma_{22}	ma_{32}		ma_{d2}
ma_{13}	ma_{23}	ma_{33}		ma_{d3}
ma_{14}	ma_{24}	ma_{34}		ma_{d4}
.	.	.		.
ma_{1n}	ma_{2n}	ma_{3n}		ma_{dn}

Projected Data on PCs

PC_1	PC_2	PC_k		PC_d
x_{11}	x_{21}	x_{k1}	….	x_{d1}
x_{12}	x_{22}	x_{k2}		x_{d2}
x_{13}	x_{23}	x_{k3}		x_{d3}
x_{14}	x_{24}	x_{k4}		x_{d4}
.	.	.		.
x_{1n}	x_{2n}	x_{kn}		x_{dn}

Dimension Reduction

Dataset

y_1	y_2	y_3		y_d
y_{11}	y_{21}	y_{31}	….	y_{d1}
y_{12}	y_{22}	y_{32}		y_{d2}
y_{13}	y_{23}	y_{33}		y_{d3}
y_{14}	y_{24}	y_{34}		y_{d4}
.	.	.		.
y_{1n}	y_{2n}	y_{3n}		y_{dn}

Mean adjusted

$y_1 - m_1$	$y_2 - m_2$	$y_3 - m_3$		$y_d - m_d$
ma_{11}	ma_{21}	ma_{31}	….	ma_{d1}
ma_{12}	ma_{22}	ma_{32}		ma_{d2}
ma_{13}	ma_{23}	ma_{33}		ma_{d3}
ma_{14}	ma_{24}	ma_{34}		ma_{d4}
.	.	.		.
ma_{1n}	ma_{2n}	ma_{3n}		ma_{dn}

Projected Data on PCs

PC_1	PC_2	PC_k		PC_d
x_{11}	x_{21}	x_{k1}	….	x_{d1}
x_{12}	x_{22}	x_{k2}		x_{d2}
x_{13}	x_{23}	x_{k3}		x_{d3}
x_{14}	x_{24}	x_{k4}		x_{d4}
.	.	.		.
x_{1n}	x_{2n}	x_{kn}		x_{dn}

k

d

K<=d

PCA Theoretical Aspects

Let D be the dataset with n number of samples and each sample is defined by a d dimensional features.

y_1	y_2	y_3		y_d
y_{11}	y_{21}	y_{31}	....	y_{d1}
y_{12}	y_{22}	y_{32}		y_{d2}
y_{13}	y_{23}	y_{33}		y_{d3}
y_{14}	y_{24}	y_{34}		y_{d4}
.	.	.		.
y_{1n}	y_{2n}	y_{3n}		y_{dn}

D

$\overline{v}$ unit vector

x1
x2
x3
xd

Projection of D on vector $\overline{v}$: $D.\overline{v}$

Best fit vector $\overline{v}$,
vector with maximum variance: $\max_{\overline{v}} var(D.\overline{v})$

Variance of the Projected Points on $\bar{v}$

Let D be the dataset with n number of samples and each sample is defined by a d dimensional features.

y_1	y_2	y_3		y_d
y_{11}	y_{21}	y_{31}	….	y_{d1}
y_{12}	y_{22}	y_{32}		y_{d2}
y_{13}	y_{23}	y_{33}		y_{d3}
y_{14}	y_{24}	y_{34}		y_{d4}
.	.	.		.
y_{1n}	y_{2n}	y_{3n}		y_{dn}

D

$\bar{v}$ Unit vector

$\bar{v}$
x_1
x_2
x_3
x_d

$$var(D\overline{v}) = \bar{v}^T S \bar{v}$$

Where *S* is the covariance matrix of D

Variance of the Projected Points on $\bar{v}$

Let D be the dataset with n number of samples and each sample is defined by a d dimensional features.

y_1	y_2	y_3		y_d
y_{11}	y_{21}	y_{31}	….	y_{d1}
y_{12}	y_{22}	y_{32}		y_{d2}
y_{13}	y_{23}	y_{33}		y_{d3}
y_{14}	y_{24}	y_{34}		y_{d4}
.	.	.		.
y_{1n}	y_{2n}	y_{3n}		y_{dn}

$\bar{v}$
x_1
x_2
x_3
x_d

D$\bar{v}$
p_1
p_2
p_3
p_n

$$var(D.v) = \boldsymbol{v}^T S \boldsymbol{v}$$

Where *S* is the covariance matrix of D

Variance of the projected points

$$var(D.\boldsymbol{v}) = \frac{1}{n-1}\sum_{1}^{n}(p_i - \bar{p})^2$$

Where $\bar{p}$ is the projected sample mean

Variance of the Projected Points on $\bar{v}$

Let D be the dataset with n number of samples and each sample is defined by a d dimensional features.

y_1	y_2	y_3		y_d	$\bar{v}$	$D\bar{v}$
y_{11}	y_{21}	y_{31}	....	y_{d1}	x_1	p_1
y_{12}	y_{22}	y_{32}		y_{d2}	x_2	p_2
y_{13}	y_{23}	y_{33}		y_{d3}	x_3	p_3
y_{14}	y_{24}	y_{34}		y_{d4}		
.	.	.		.	x_d	
y_{1n}	y_{2n}	y_{3n}		y_{dn}		p_n

$$var(D.v) = \boldsymbol{v}^T S \boldsymbol{v}$$

Where S is the covariance matrix of D

Variance of the projected points

$$var(D.\boldsymbol{v}) = \frac{1}{n-1}\sum_{1}^{n}(p_i - \bar{p})^2$$

Where $\bar{p}$ is the projected sample mean

$$var(D.\boldsymbol{v}) = \frac{1}{n-1}\sum_{1}^{n}(\boldsymbol{v}^T\boldsymbol{d}_i - \boldsymbol{v}^T\bar{\boldsymbol{d}})^2$$

Where $\bar{\boldsymbol{d}}$ is the sample mean

Variance of the Projected Points on $\overline{v}$

Let D be the dataset with n number of samples and each sample is defined by a d dimensional features.

y_1	y_2	y_3		y_d	$\overline{v}$	$D\overline{v}$
y_{11}	y_{21}	y_{31}	....	y_{d1}	x_1	p_1
y_{12}	y_{22}	y_{32}		y_{d2}	x_2	p_2
y_{13}	y_{23}	y_{33}		y_{d3}	x_3	p_3
y_{14}	y_{24}	y_{34}		y_{d4}		
.	.	.		.	x_d	
y_{1n}	y_{2n}	y_{3n}		y_{dn}		p_n

$$var(D.v) = \boldsymbol{v}^T S \boldsymbol{v}$$

Where S is the covariance matrix of D

Variance of the projected points

$$var(D.\boldsymbol{v}) = \frac{1}{n-1}\sum_{1}^{n}(p_i - \bar{p})^2$$

Where $\bar{p}$ is the projected sample mean

$$var(D.\boldsymbol{v}) = \frac{1}{n-1}\sum_{1}^{n}(\boldsymbol{v}^T \boldsymbol{d}_i - \boldsymbol{v}^T \overline{\boldsymbol{d}})^2$$

Where $\overline{\boldsymbol{d}}$ is the sample mean

$$var(D.v) = \boldsymbol{v}^T S \boldsymbol{v}$$ where S is the covariance matric of D i.e. $$S = \frac{1}{n-1}\sum_{1}^{n}(\boldsymbol{d}_i - \overline{\boldsymbol{d}})(\boldsymbol{d}_i - \overline{\boldsymbol{d}})^T$$

Best vector that maximizes the variance

$$var(D.v) = \boldsymbol{v}^T S \boldsymbol{v}$$

Optimization

$$\max_{\boldsymbol{v}}(\boldsymbol{v}^T S \boldsymbol{v})$$

Best vector that maximizes the variance

$$var(D.v) = \boldsymbol{v}^T S \boldsymbol{v}$$

Optimization

$$\max_{\boldsymbol{v}}(\boldsymbol{v}^T S \boldsymbol{v})$$

It is quadratic and has no upper bound

Best vector that maximizes the variance

$$var(D.v) = \boldsymbol{v}^T S \boldsymbol{v}$$

Optimization

$\max_{\boldsymbol{v}} (\boldsymbol{v}^T S \boldsymbol{v})$ It is quadratic and has no upper bound

Maximize with constraints

$\boldsymbol{v}^T S \boldsymbol{v}$ s.t. $\boldsymbol{v}^T \boldsymbol{v} = 1$

Best vector that maximizes the variance

$$var(D.v) = \boldsymbol{v}^T S\boldsymbol{v}$$

Optimization

$$\max_{\boldsymbol{v}}(\boldsymbol{v}^T S\boldsymbol{v})$$

It is quadratic and has no upper bound

Maximize

$\boldsymbol{v}^T S\boldsymbol{v} - \boldsymbol{\lambda}(\boldsymbol{v}^T\boldsymbol{v} - \mathbf{1})$ **where λ Lagrange Multiplier**

Best vector that maximizes the variance

$$L(\boldsymbol{v}, \lambda) = \boldsymbol{v}^T S \boldsymbol{v} - \boldsymbol{\lambda}(\boldsymbol{v}^{\boldsymbol{T}} \boldsymbol{v} - \mathbf{1})$$

$$\frac{\partial L}{\partial \boldsymbol{v}} = 2S\boldsymbol{v} - 2\lambda\boldsymbol{v} = 0$$

$$S\boldsymbol{v} = \lambda\boldsymbol{v}$$

Best vector that maximizes the variance

$$L(\boldsymbol{v}, \lambda) = \boldsymbol{v}^T S \boldsymbol{v} - \boldsymbol{\lambda}(\boldsymbol{v}^{\boldsymbol{T}} \boldsymbol{v} - \mathbf{1})$$

$$\frac{\partial L}{\partial \boldsymbol{v}} = 2S\boldsymbol{v} - 2\lambda\boldsymbol{v} = 0$$

$$\Rightarrow S\boldsymbol{v} = \lambda\boldsymbol{v}$$ ← **Eigenvector expression**

Best vector that maximizes the variance

$$L(\boldsymbol{v}, \lambda) = \boldsymbol{v}^T S \boldsymbol{v} - \boldsymbol{\lambda}(\boldsymbol{v}^T \boldsymbol{v} - \mathbf{1})$$

$$\frac{\partial L}{\partial \boldsymbol{v}} = 2S\boldsymbol{v} - 2\lambda\boldsymbol{v} = 0$$

$$\Rightarrow S\boldsymbol{v} = \lambda\boldsymbol{v}$$ ← **Eigenvector expression**

Which Eigenvector ?

Best vector that maximizes the variance

$$L(\boldsymbol{v}, \lambda) = \boldsymbol{v}^T S \boldsymbol{v} - \boldsymbol{\lambda}(\boldsymbol{v}^T \boldsymbol{v} - \mathbf{1})$$

$$\frac{\partial L}{\partial \boldsymbol{v}} = 2S\boldsymbol{v} - 2\lambda\boldsymbol{v} = 0$$

$$\Rightarrow S\boldsymbol{v} = \lambda\boldsymbol{v}$$

$$\Rightarrow \boldsymbol{v}^T S\boldsymbol{v} = \lambda$$

$$\Rightarrow var(D\boldsymbol{v}) = \boldsymbol{v}^T S\boldsymbol{v} = \lambda$$ ← **Largest Eigenvalue**

Principal Eigenvector

Dimension Reduction

- 1^{st} Principal Component => 1^{st} Principal Eigenvector
- 2^{nd} Principal Component => vector perpendicular to 1^{st} PC
  => 2^{nd} Principal Eigenvector
- 3^{rd} Principal Component => vector perpendicular to 1^{st} PC and 2^{nd} PC
  => 2^{nd} Principal Eigenvector

…so on

Select the k principal components and project the data point over the selected eigenvectors

Dimension Reduction

Select the k principal components and project the data point over the selected eigenvectors

U be the matrix whose column vectors are the selected eigenvectors of D such that

1st column vector is the 1 principal eigenvector, 2nd is the 2nd eigenvector ….

$$\acute{D} = DU$$

$$\underset{\acute{D}\ (n\times k)}{\begin{array}{cccc} EV_1 & EV_2 & & EV_k \\ \hline p_{11} & p_{21} & \cdots & p_{k1} \\ p_{12} & p_{22} & & p_{k2} \\ p_{13} & p_{23} & & p_{k3} \\ p_{14} & p_{24} & & p_{k4} \\ . & . & & . \\ p_{1n} & p_{2n} & & p_{kn} \end{array}} = \underset{D\ (n\times d)}{\begin{array}{ccccc} y_1 & y_2 & y_3 & & y_d \\ \hline y_{11} & y_{21} & y_{31} & \cdots & y_{d1} \\ y_{12} & y_{22} & y_{32} & & y_{d2} \\ y_{13} & y_{23} & y_{33} & & y_{d3} \\ y_{14} & y_{24} & y_{34} & & y_{d4} \\ . & . & . & & . \\ y_{1n} & y_{2n} & y_{3n} & & y_{dn} \end{array}} \underset{U\ (d\times k)}{\begin{array}{cccc} EV_1 & EV_2 & & EV_k \\ \hline x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & & x_{k2} \\ x_{13} & x_{23} & & x_{k3} \\ . & . & & . \\ x_{1d} & x_{2d} & & x_{kd} \end{array}}$$

PCA Summary

- Covariance Matrix S of the Data matrix D
- Estimate Eigenvectors of the S
- Select k principal eigenvectors
- Project the data matrix on the selected k Eigenvectors