Geometric Interpretation of Linear Regression

Nipun Batra

February 7, 2023

IIT Gandhinagar

Linear Combination of Vectors

Let $v_1, v_2, v_3, \ldots, v_i$ be vectors in $\mathbb{R}^D$, where D denotes the dimensions.

Linear Combination of Vectors

Let $v_1, v_2, v_3, \ldots, v_i$ be vectors in $\mathbb{R}^D$, where D denotes the dimensions.

A linear combination of $v_1, v_2, v_3, \ldots, v_i$ is of the following form

Linear Combination of Vectors

Let $v_1, v_2, v_3, \ldots, v_i$ be vectors in $\mathbb{R}^D$, where D denotes the dimensions.
A linear combination of $v_1, v_2, v_3, \ldots, v_i$ is of the following form

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \cdots + \alpha_i v_i$$

where $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_i \in \mathbb{R}$

Span of vectors

Let $v_1, v_2, \ldots, v_i$ be vectors in $\mathbb{R}^D$, with D dimensions.

Span of vectors

Let $v_1, v_2, \ldots, v_i$ be vectors in $\mathbb{R}^D$, with D dimensions.
The span of $v_1, v_2, \ldots, v_i$ is denoted by $\text{SPAN}\{v_1, v_2, \ldots, v_i\}$

Span of vectors

Let $v_1, v_2, \ldots, v_i$ be vectors in $\mathbb{R}^D$, with D dimensions.
The span of $v_1, v_2, \ldots, v_i$ is denoted by SPAN$\{v_1, v_2, \ldots, v_i\}$

$$\{\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_i v_i \quad | \quad \alpha_1, \alpha_2, \ldots, \alpha_i \in \mathbb{R}\}$$

Span of vectors

Let $v_1, v_2, \ldots, v_i$ be vectors in $\mathbb{R}^D$, with D dimensions.
The span of $v_1, v_2, \ldots, v_i$ is denoted by $\mathrm{SPAN}\{v_1, v_2, \ldots, v_i\}$

$$\{\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_i v_i \quad | \quad \alpha_1, \alpha_2, \ldots, \alpha_i \in \mathbb{R}\}$$

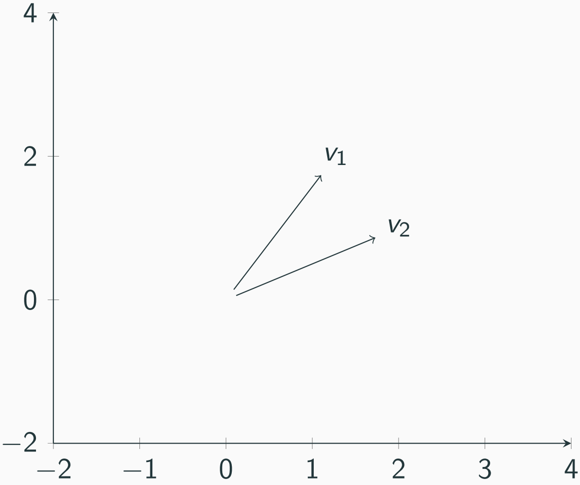
It is the set of all vectors that can be generated by linear combinations of $v_1, v_2, \ldots, v_i$.

Example

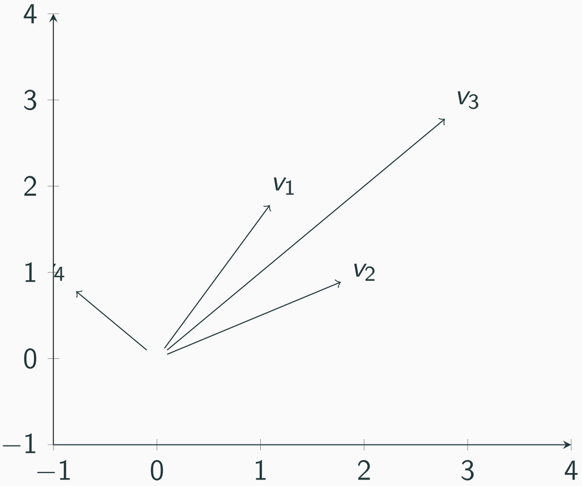
Find the span of $\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right)$

Example

Find the span of ($\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$)



Example



$v_3 = v_1 + v_2$ and $v_4 = v_1 - v_2$

$\text{Span}((v_1, v_2)) \in \mathcal{R}^2$

Example

Find the span of $\left(\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\4\end{bmatrix}\right)$

Example

Find the span of $(\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\4\end{bmatrix})$

Can we obtain a point (x, y) s.t. $x = 3y$?

Example

Find the span of $\left(\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\4\end{bmatrix}\right)$

Can we obtain a point (x, y) s.t. $x = 3y$?

No

Example

Find the span of $\left(\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\4\end{bmatrix}\right)$

Can we obtain a point (x, y) s.t. $x = 3y$?

No

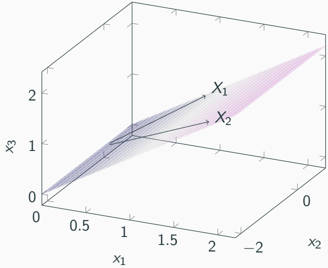
Span of the above set is along the line $y = 2x$

Example

Find the span of $\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix} \right)$

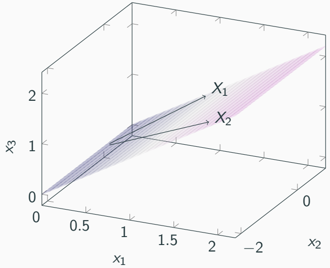
Example

Find the span of ($\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix}$)



Example

Find the span of ($\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix}$)



The span is the plane $z = x$ or $x_3 = x_1$

Geometric Interpretation

Consider X and y as follows.

$$\mathbf{X} = \begin{pmatrix} 1 & 2 \\ 1 & -2 \\ 1 & 2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 8.8957 \\ 0.6130 \\ 1.7761 \end{pmatrix}$$

- We are trying to learn θ for $\hat{y} = X\theta$ such that $||y - \hat{y}||_2$ is minimised

Geometric Interpretation

Consider X and y as follows.

$$\mathbf{X} = \begin{pmatrix} 1 & 2 \\ 1 & -2 \\ 1 & 2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 8.8957 \\ 0.6130 \\ 1.7761 \end{pmatrix}$$

- We are trying to learn θ for $\hat{y} = X\theta$ such that $||y - \hat{y}||_2$ is minimised
- Consider the two columns of X. Can we write $X\theta$ as the span of $\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix}\right)$?

Geometric Interpretation

Consider X and y as follows.

$$\mathbf{X} = \begin{pmatrix} 1 & 2 \\ 1 & -2 \\ 1 & 2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 8.8957 \\ 0.6130 \\ 1.7761 \end{pmatrix}$$

- We are trying to learn θ for $\hat{y} = X\theta$ such that $||y - \hat{y}||_2$ is minimised
- Consider the two columns of X. Can we write $X\theta$ as the span of $(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix})$?
- We wish to find $\hat{y}$ such that

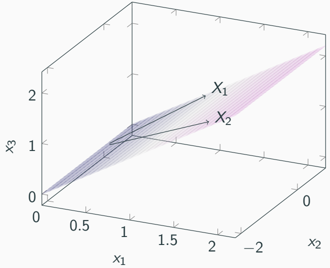
$$\underset{\hat{y} \in SPAN\{\bar{x_1}, \bar{x_2}, \ldots, \bar{x_D}\}}{\arg\min} ||y - \hat{y}||_2$$

Geometric Interpretation

$$\text{Span of } \left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix} \right)$$

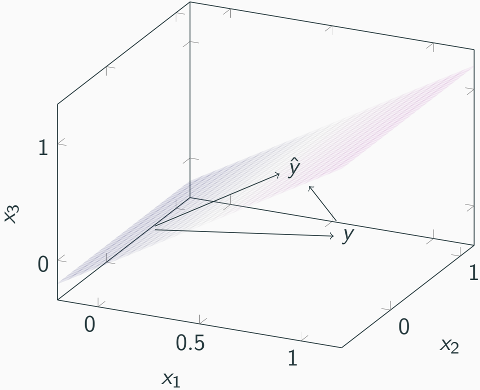
Geometric Interpretation

Span of $\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix} \right)$



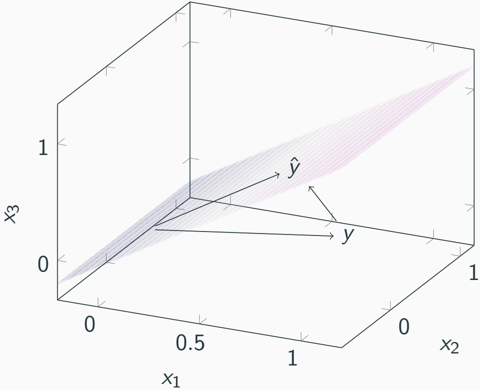
The span is the plane $z = x$ or $x_3 = x_1$

Geometric Interpretation



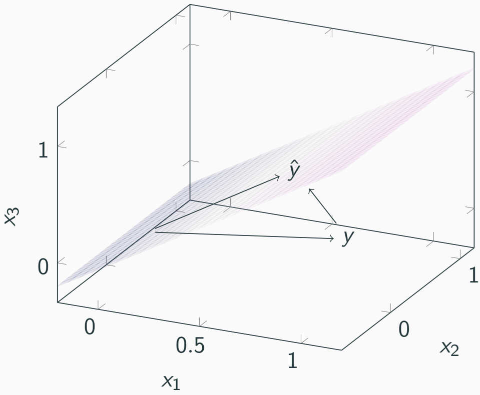
- We seek a $\hat{y}$ in the span of the columns of X such that it is closest to y

Geometric Interpretation



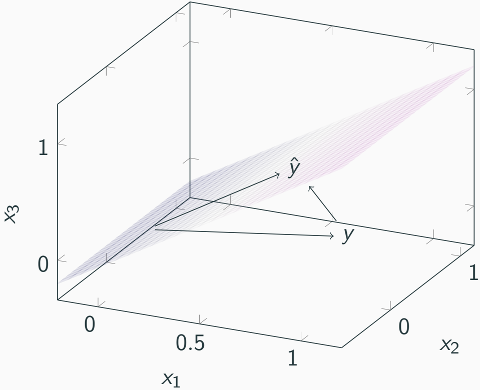
- We seek a $\hat{y}$ in the span of the columns of X such that it is closest to y
- This happens when $y - \hat{y} \perp x_j \forall j$ or $x_j^T(y - \hat{y}) = 0$

Geometric Interpretation



- We seek a $\hat{y}$ in the span of the columns of X such that it is closest to y
- This happens when $y - \hat{y} \perp x_j \forall j$ or $x_j^T(y - \hat{y}) = 0$
- $X^T(y - X\theta) = 0$

Geometric Interpretation



- We seek a $\hat{y}$ in the span of the columns of X such that it is closest to y
- This happens when $y - \hat{y} \perp x_j \forall j$ or $x_j^T(y - \hat{y}) = 0$
- $X^T(y - X\theta) = 0$
- $X^T y = X^T X\theta$ or $\hat{\theta} = (X^T X)^{-1} X^T y$