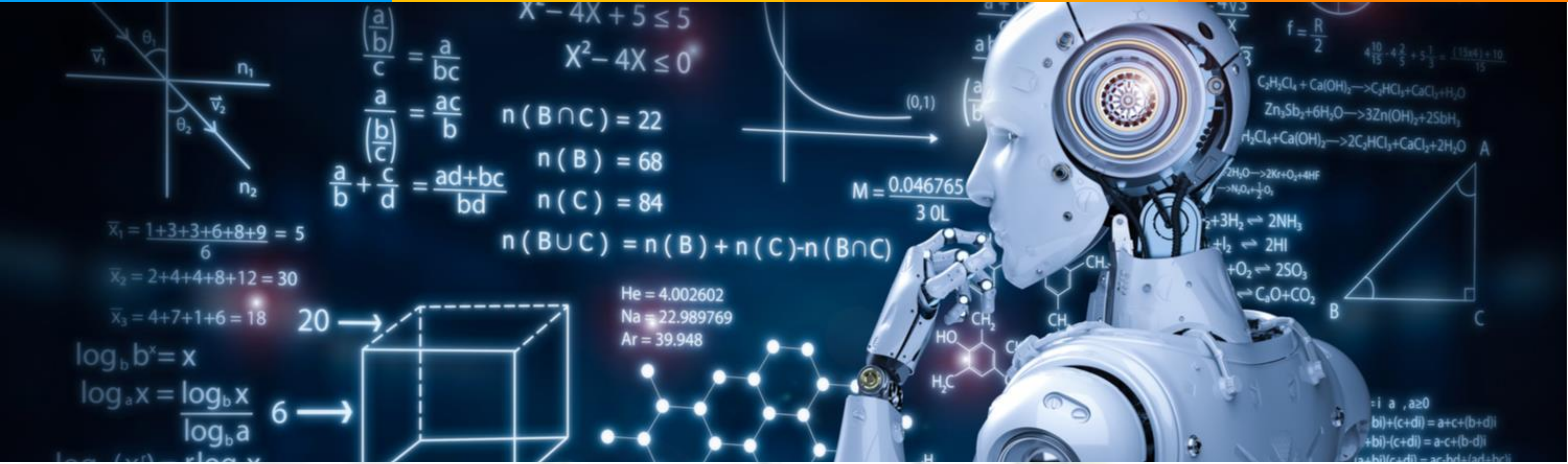
Duration: 60 min | Srikanth Shenoy | AML 5101 | M.E. AI & ML

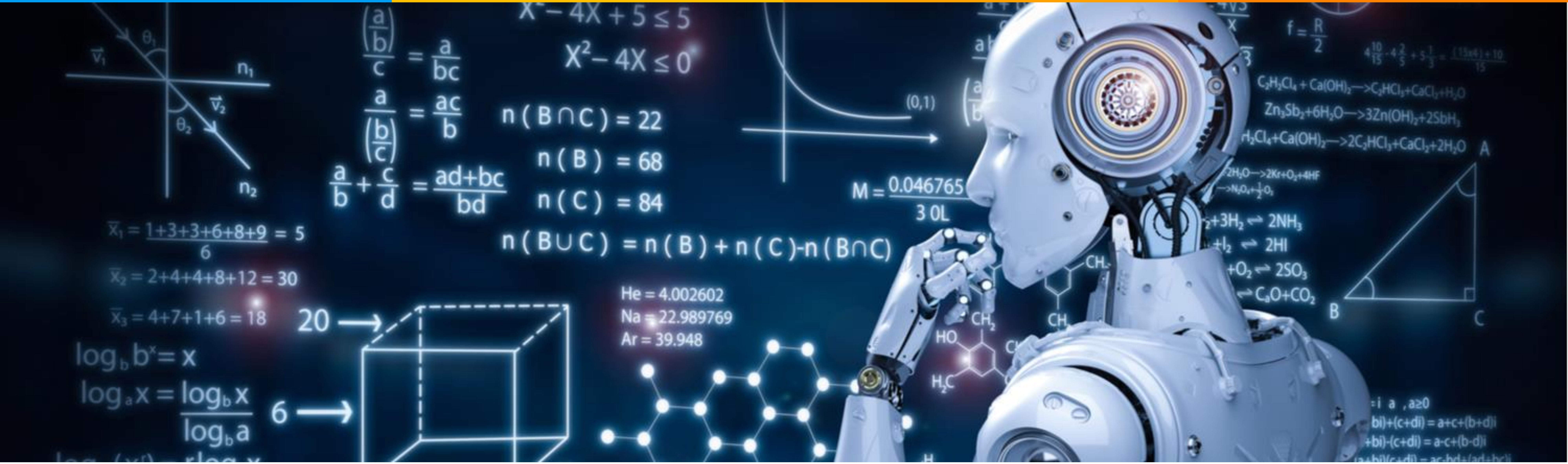


Lecture 08 Linear Combinations, & Independence

Recap

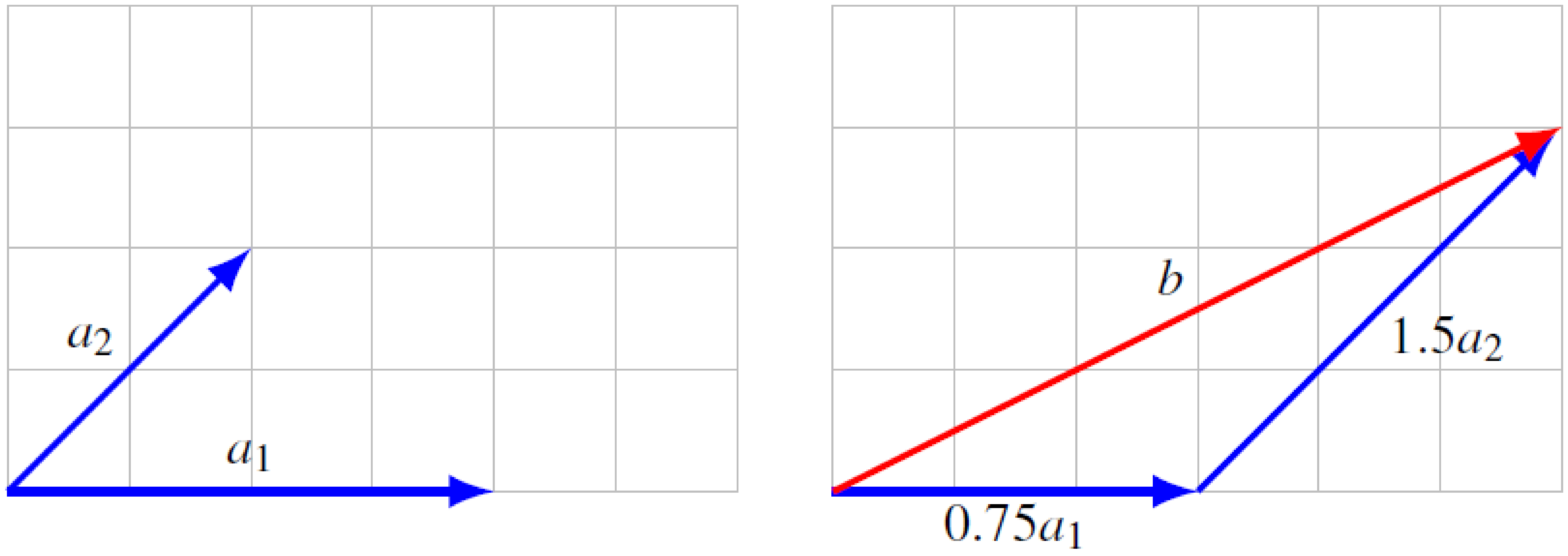
- Correlation, auto correlation

Duration: 60 min | Srikanth Shenoy | AML 5101 | M.E. AI & ML



Linear Combinations

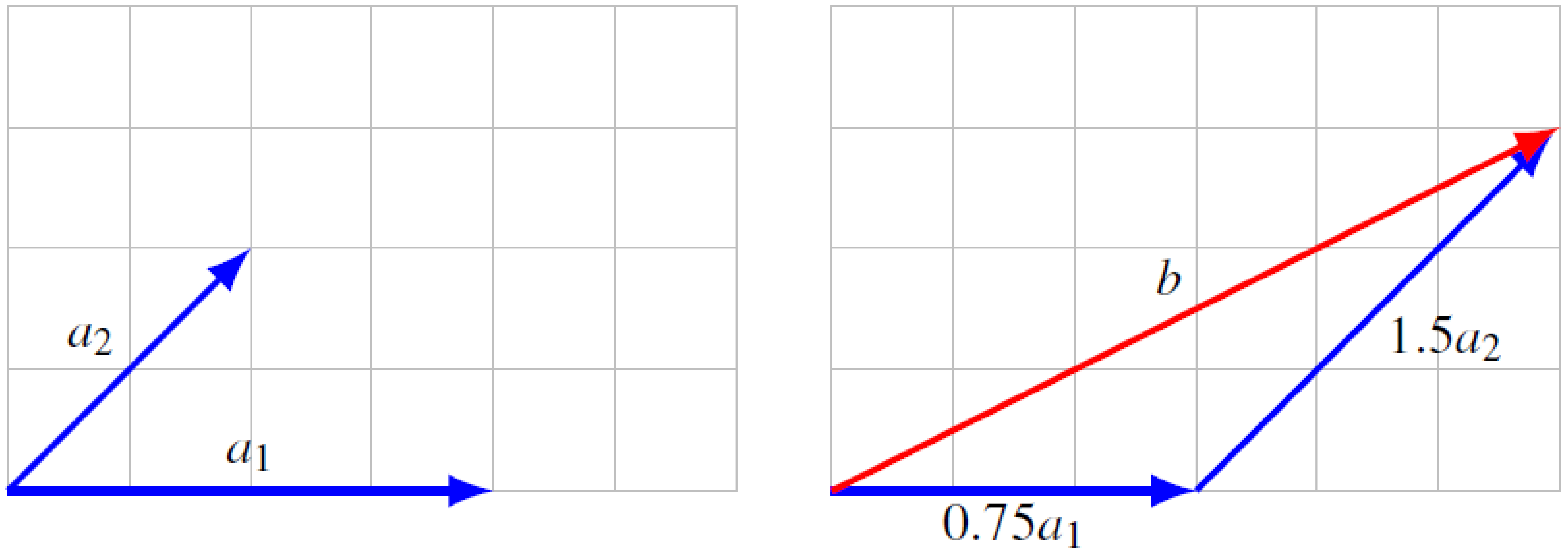
Linear Combinations



- Scale and add vectors

Linear Combinations

- Definition: $\beta_1\mathbf{x_1} + \beta_2\mathbf{x_2} + \ldots + \beta_n\mathbf{x_n}$
  - $\beta_1, \beta_2, \ldots, \beta_n$ are scalars
  - $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$ are vectors
- Simply put: Scale and add vectors



Linear combination examples – Audio Mixing

- Sound technician at music concert gets sound inputs
- Every input is a vector over time window t to t+n-1
- s from saxophone
- g from guitar
- v from vocal

$$s = \begin{bmatrix} s_1 \\ s_2 \\ .. \\ s_n \end{bmatrix} \quad g = \begin{bmatrix} g_1 \\ g_2 \\ .. \\ g_n \end{bmatrix} \quad v = \begin{bmatrix} v_1 \\ v_2 \\ .. \\ v_n \end{bmatrix}$$

$$a = \beta_1 \mathbf{s} + \beta_2 \mathbf{g} + \beta_3 \mathbf{v}$$

$$a = \beta_1 \begin{bmatrix} s_1 \\ s_2 \\ .. \\ s_n \end{bmatrix} + \beta_2 \begin{bmatrix} g_1 \\ g_2 \\ .. \\ g_n \end{bmatrix} + \beta_3 \begin{bmatrix} v_1 \\ v_2 \\ .. \\ v_n \end{bmatrix}$$

Vector as a linear combination of unit vectors

$$b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$b = b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + b_3\mathbf{e}_3$$

$$\mathbf{b} = b_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + b_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Matrix vector product is linear combination

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

$$X = \begin{bmatrix} 76 & 126 & 38 \\ 74 & 120 & 38 \\ 72 & 118 & 37.5 \\ 78 & 136 & 37 \end{bmatrix} \qquad v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$v_1 \mathbf{HR} + v_2 \mathbf{BP} + v_3 \mathbf{Temp}$$

Check this with your known formula for matrix multiplication

$$Xv = v_1 \begin{bmatrix} 76 \\ 74 \\ 72 \\ 78 \end{bmatrix} + v_2 \begin{bmatrix} 126 \\ 120 \\ 118 \\ 136 \end{bmatrix} + v_3 \begin{bmatrix} 38 \\ 38 \\ 37.5 \\ 37 \end{bmatrix}$$

PCA is all about linear combination

What is the intuitive meaning of adding fractions of heterogeneous feature vectors?

	HR	BP	Temp
Patient-1	76	126	38.0
Patient-2	74	120	38.0
Patient-3	72	118	37.5
Patient-4	78	136	37.0

New synthetic feature

Replaces original features

Called Principal Components PC1, PC2 etc.

$$\beta_1 \mathbf{HR} + \beta_2 \mathbf{BP} + \beta_3 \mathbf{Temp}$$

PCA Goals

1. **Feature Extraction**
2. **Dimensionality Reduction**

Fractions to mix features are carefully calculated

PCA – Creating synthetic features PC1, PC2…

Name	Diastolic BP	Systolic BP
Patient1	78.00	126.00
Patient2	80.00	128.00
Patient3	81.00	127.00
Patient4	82.00	130.00
Patient5	84.00	130.00
Patient6	86.00	132.00
Variance	8.17	4.97

- Variance of DBP is roughly double of SBP (but not quite)

$$\beta_1 = 0.8$$
$$\beta_2 = 0.6$$

$$\beta_1^2 \approx 2 \times \beta_2^2$$

- Create synthetic feature PC1 by linear combination

$$PC1 = \beta_1\mathbf{DBP} + \beta_2\mathbf{SBP}$$

PCA is all about explained variance

Name	Diastolic BP	Systolic BP	PC1
Patient1	78.00	126.00	138.00
Patient2	80.00	128.00	140.80
Patient3	81.00	127.00	141.00
Patient4	82.00	130.00	143.60
Patient5	84.00	130.00	145.20
Patient6	86.00	132.00	148.00
Variance	8.17	4.97	12.74

$$PC1 = \beta_1 \mathbf{DBP} + \beta_2 \mathbf{SBP}$$

$$\beta_1 = 0.8$$

$$\beta_2 = 0.6$$

- Total Variance = 13.14
- PC1 Variance = 12.74

$$\frac{12.74}{13.14} \times 100 = 97 percent$$

Linear combination examples – PCA

$$PC1 = \beta_1 \mathbf{DBP} + \beta_2 \mathbf{SBP}$$

Beta1	Beta2	PC1 Variance
0.8	0.6	12.74
0.6	0.8	11.8
0.98	0.2	10.4
0.2	0.98	7.4

- Beta1 = 0.8, Beta=0.6 fraction making PC1 captures **MAXIMUM** variance

Dimensionality reduction with PCA

$$X = \begin{bmatrix} 76 & 126 & 38 & .... & x_n^{(1)} \\ 74 & 120 & 38 & .... & x_n^{(2)} \\ .. & .. & .. & .. & .. \\ 72 & 118 & 37.5 & .... & x_n^{(i)} \\ .. & .. & .. & .. & .. \\ 78 & 136 & 37 & .... & x_m^{(n)} \end{bmatrix}$$

$$PC1 = \alpha_1 \mathbf{x_1} + \alpha_2 \mathbf{x_2} + ... + \alpha_n \mathbf{x_n}$$

$$PC2 = \beta_1 \mathbf{x_1} + \beta_2 \mathbf{x_2} + ... + \beta_n \mathbf{x_n}$$

$$PCi = ...$$

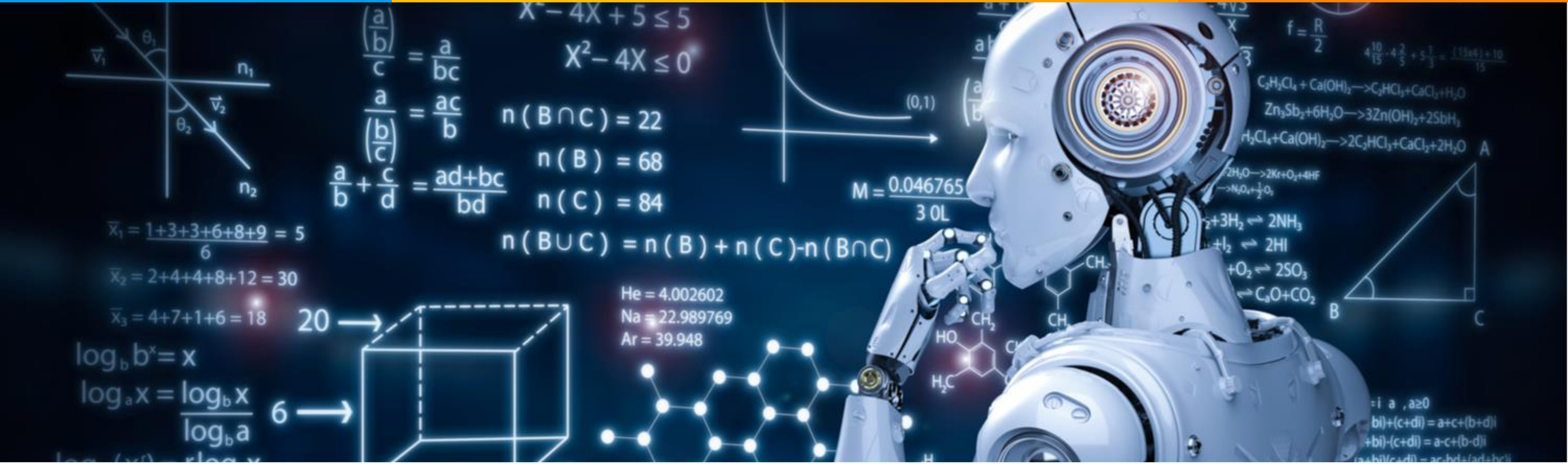
$$PCn = \eta_1 \mathbf{x_1} + \eta_2 \mathbf{x_2} + ... + \eta_n \mathbf{x_n}$$

$$\tilde{X} = \begin{bmatrix} \uparrow & \uparrow & \uparrow & \uparrow \\ PC_1 & PC_2 & .. & PC_k \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix}$$

$$k << n \quad Var(X) \approx Var(\tilde{X})$$

Current status and what’s next

- Chapter 1 & 3 are completed
- Starting chapter 5 next
- Then chapter 2 & 4
- Followed by Chapter 6
- Sessional 1 portion: Chapter 1-6



Linear Independence

Linear dependence

- Consider a set of vectors $\{\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_n}\}$
- If we can find real $\beta_1, \beta_2, \ldots \beta_n$ such that

$$\beta_1 \mathbf{x_1} + \beta_2 \mathbf{x_2} + \ldots + \beta_n \mathbf{x_n} = 0$$

- This is equivalent to saying

$$\mathbf{x_i} = \left(\frac{-\beta_1}{\beta_i}\right)\mathbf{x_1} + \left(\frac{-\beta_2}{\beta_i}\right)\mathbf{x_2} + \ldots + \left(\frac{-\beta_n}{\beta_i}\right)\mathbf{x_n}$$

- In plain English -> One vector can be expressed in terms of others

Linear independence

- $\beta_1 \mathbf{x_1} + \beta_2 \mathbf{x_2} + \ldots + \beta_n \mathbf{x_n} = 0$
- Only if $\beta_1 = \beta_2 = \ldots = \beta_n = 0$
- This is equivalent to saying
  - Any vector of the set CANNOT be expressed as linear combination of other vectors
- Standard unit vectors are linearly independent

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Linear dependence - Geometric meaning

- Geogebra demo
  - https://www.geogebra.org/calculator/eedbpvb9
  - https://www.geogebra.org/calculator/a644wefh

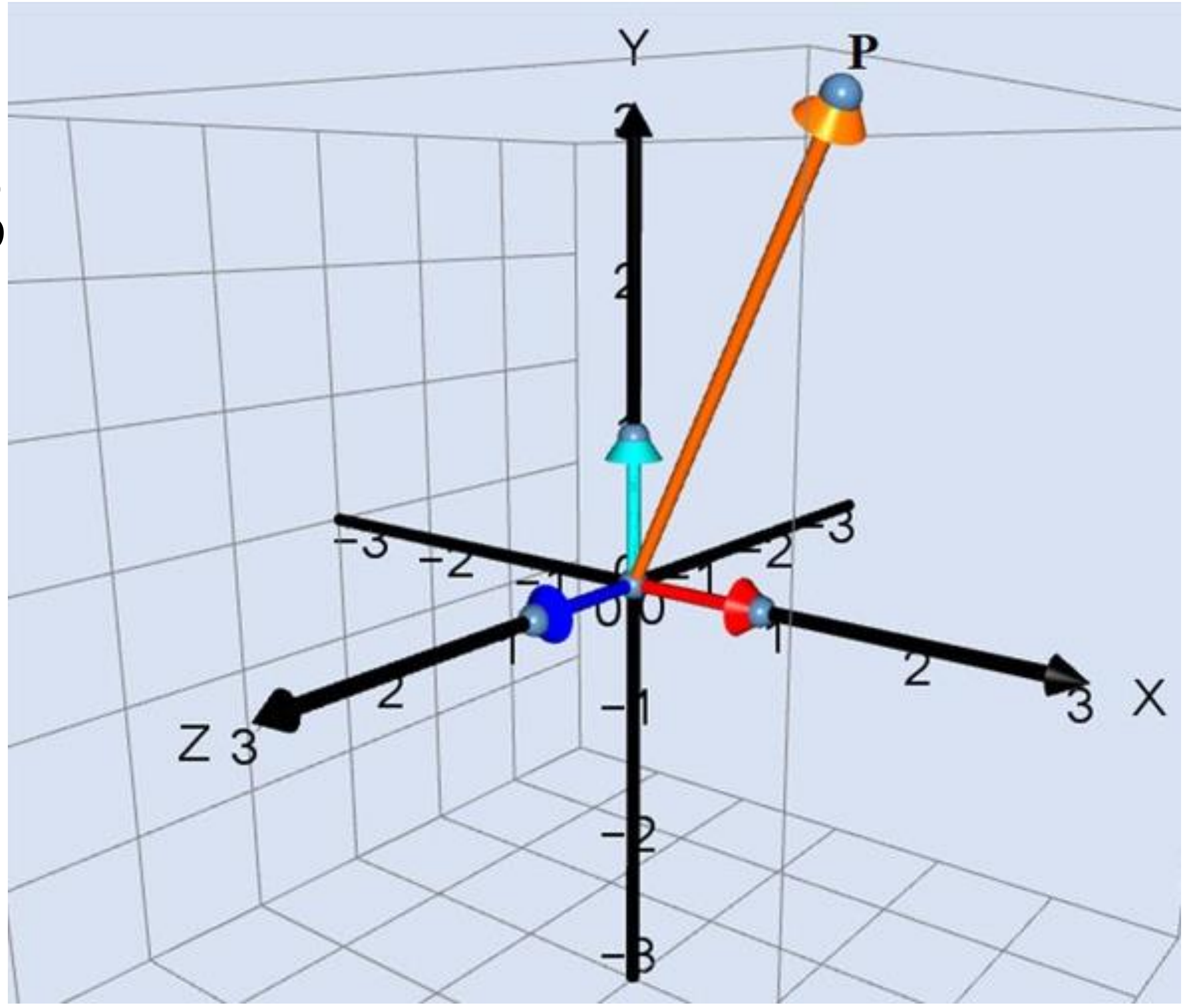
Linear dependence - Geometric meaning

- Span - How many vectors are sufficient to span 2D?
  - 2 (as long as they are not linear combo of each other)
- Can you now explain why standard unit vectors are linearly independent? $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$
- If there n+1 vectors of size n, then they have to be linearly dependent

Coefficients in linear independence

- $\mathbf{x} = \beta_1 \mathbf{x_1} + \beta_2 \mathbf{x_2} + \ldots + \beta_n \mathbf{x_n}$
- where $\{\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_n}\}$ are linearly independent
- Then $\beta_1, \beta_2, \ldots = \beta_n$ are unique
- Special case with standard unit vectors

Unit vector



$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathbf{x} = x_1 \mathbf{e_1} + x_2 \mathbf{e_2} + \ldots + x_n \mathbf{e_n}$$

- Uniqueness of coefficients is the reason why we can represent a vector with unique coefficients in vector space

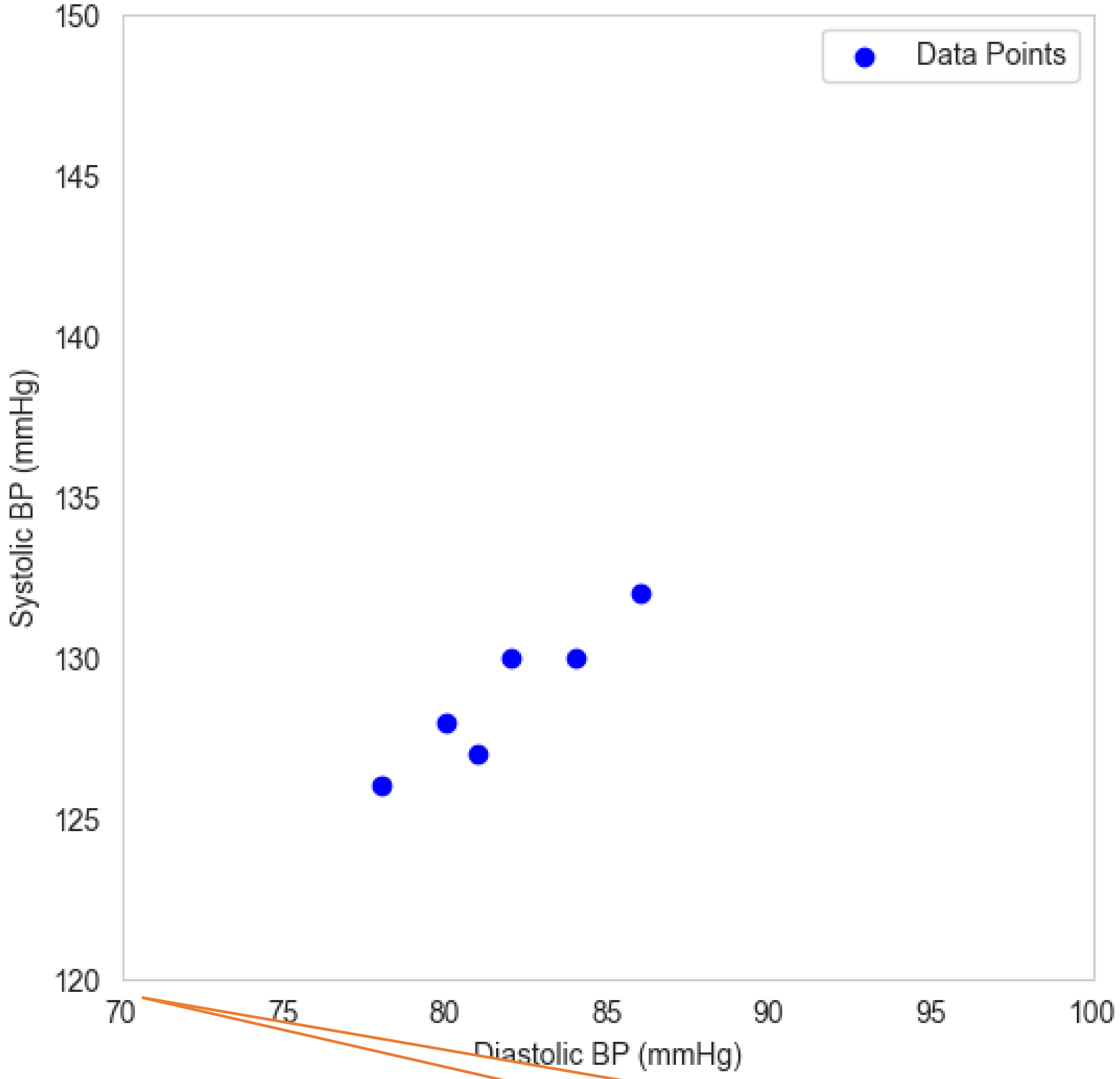
Basis

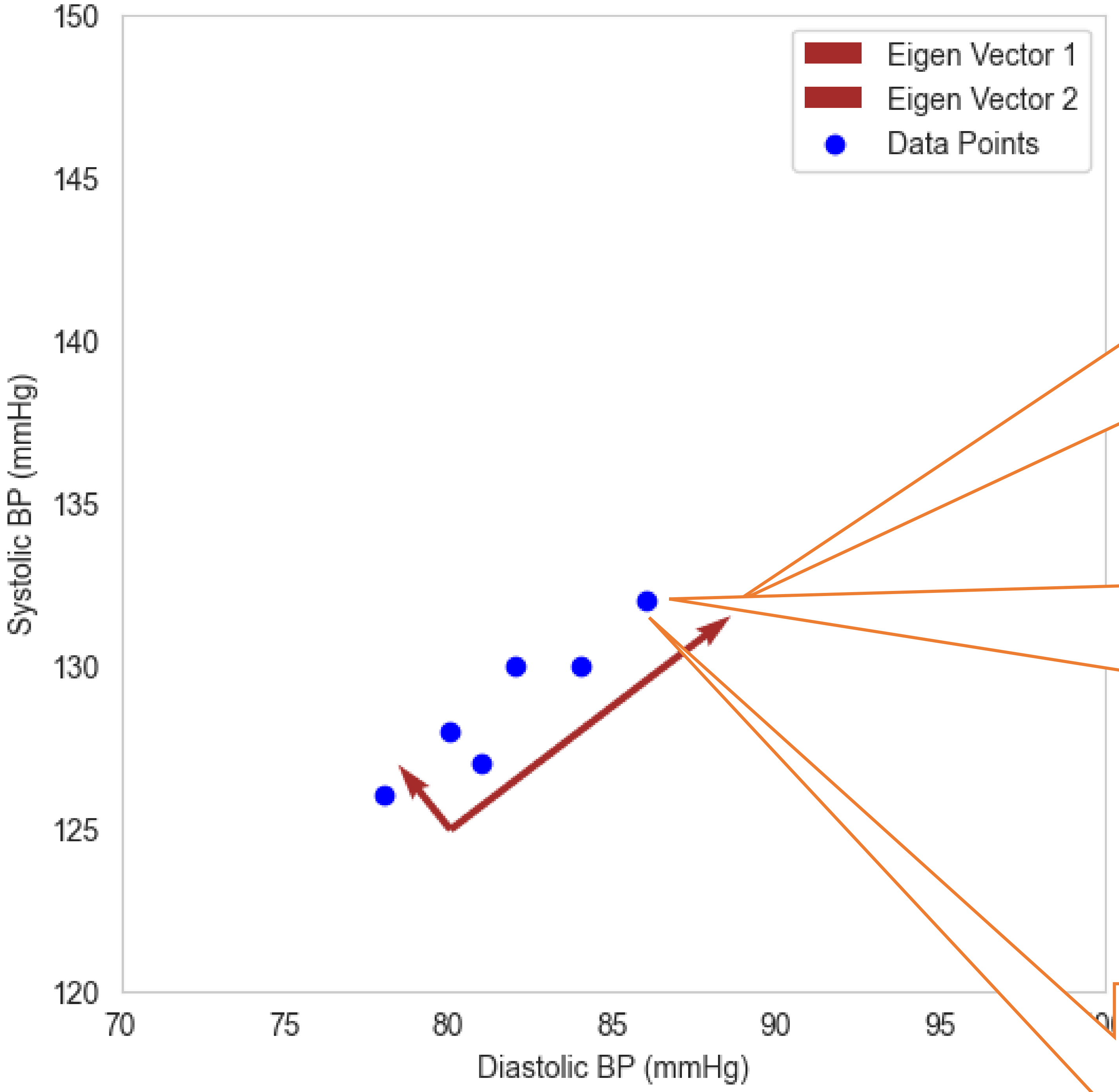
- n independent vectors $\{\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_n}\}$ basis
- Basis is frame of reference
  - E.g. coordinate system is a basis
- Standard unit vectors are special basis

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Why do Linear Algebra folks keep coming up with these names?

Basis - Why different terminology?





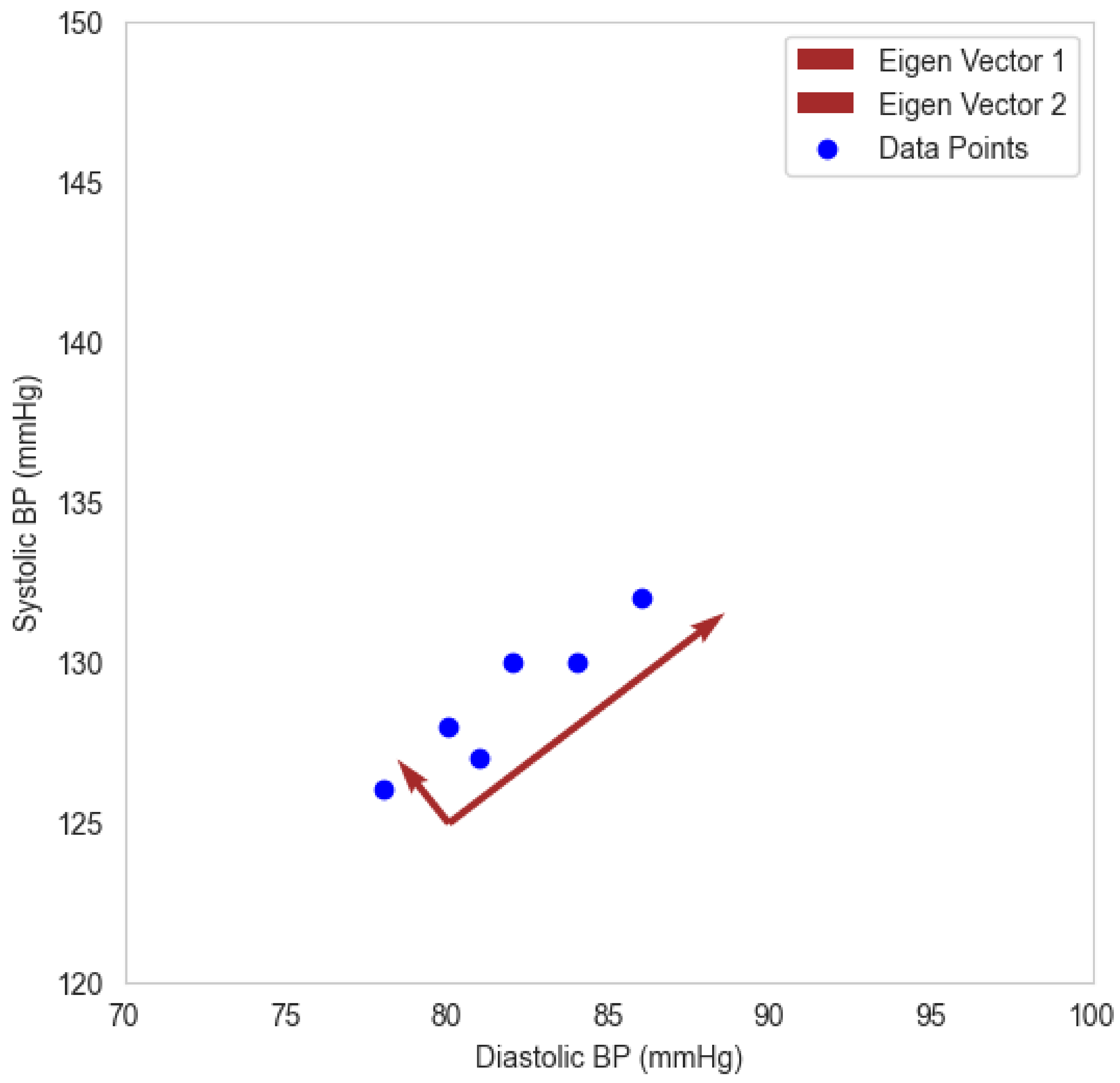
Imagine changing to this new basis

How is this point represented in new basis?

Take Matrix vector product

Standard basis

Matrix multiplication as change of basis



$$X = \begin{bmatrix} 78 & 126 \\ 80 & 128 \\ 81 & 127 \\ 82 & 130 \\ 84 & 130 \\ 86 & 132 \end{bmatrix}$$

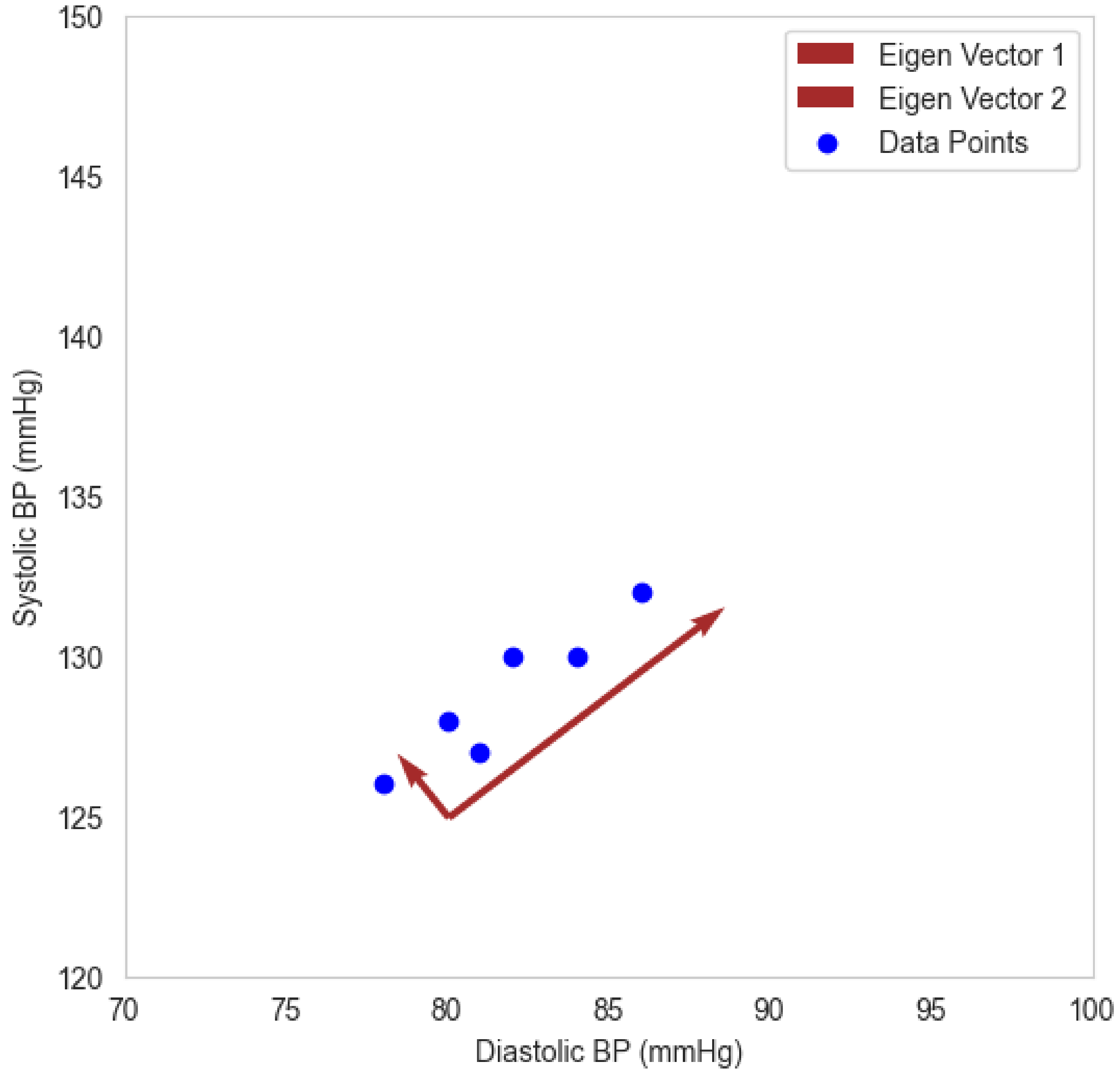
$$v_x = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad v_y = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$v_x = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad v_y = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \qquad Vx^{(1)} = \begin{bmatrix} 138.6 \\ 52.5 \end{bmatrix}$$

- What is matrix-vector product $Vx^{(1)}$

Matrix multiplication as change of basis



$$X = \begin{bmatrix} 78 & 126 \\ 80 & 128 \\ 81 & 127 \\ 82 & 130 \\ 84 & 130 \\ 86 & 132 \end{bmatrix} \quad v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad v_2 = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$Xv_1 = \begin{bmatrix} 138.6 \\ 141.4 \\ 141.6 \\ 144.2 \\ 145.8 \\ 148.6 \end{bmatrix} = PC1$$

$$Vx^{(1)} = \begin{bmatrix} 138.6 \\ 52.5 \end{bmatrix}$$

$$Xv_2 = PC2$$

How are PC1, PC2 obtained?

- Eigen decomposition of Covariance Matrix
- PC1 … PCn are product of feature matrix with Eigen vectors 1, 2 and so on
- Beta1, Beta2 used to form PC1 etc. are the entries of the Eigen vector 1.
- Eigen Vectors are normalized (unit magnitude)
- Hence

$$\beta_1^2 + \beta_2^2 + \ldots.\beta_n^2 = 1$$

PCA for dimensionality reduction

- Explain maximum possible variance of original data with as PC1, PC2… PCn
- PC1, PC2 etc.. are linear combination of features
- Adding vectors of different data type
- Adding patient BP, HR, & Temp is normally meaningless
- But here it has great mathematical utility
- **Takeaway: Principal components obtained from PCA are linear combinations of features so that max feature variance is explained with as less PCs**


QUESTIONS


Thank You!