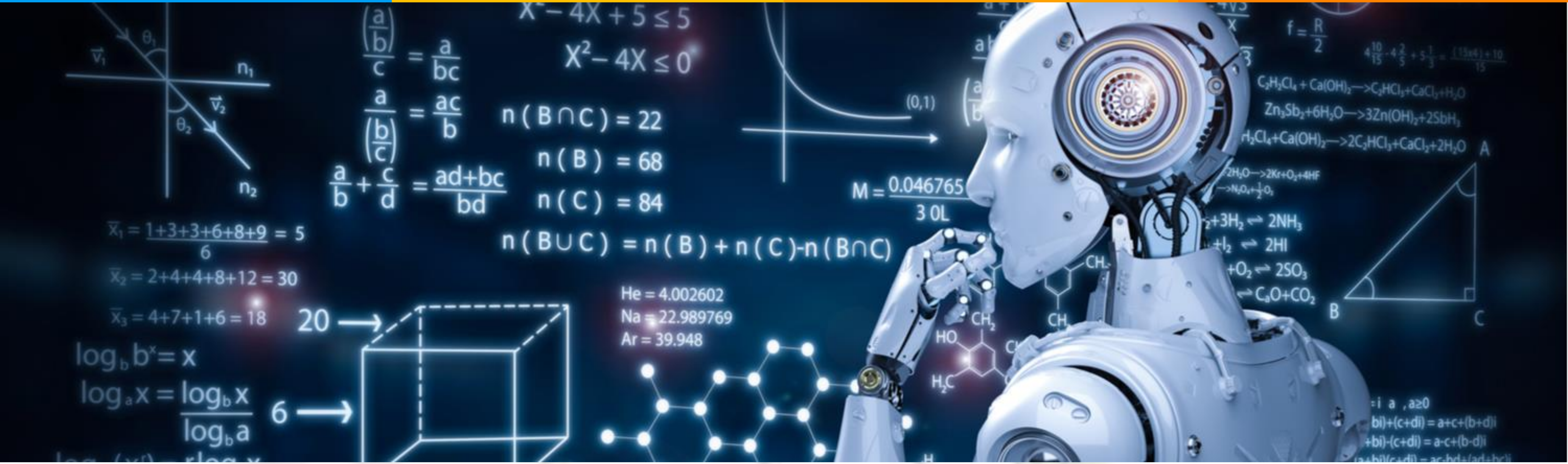
Duration: 60 min | Srikanth Shenoy | AML 5102 | M.E. AI & ML



Lecture 33 PCA & More

Matrix Vector Product

Matrix Vector multiplication as transformation

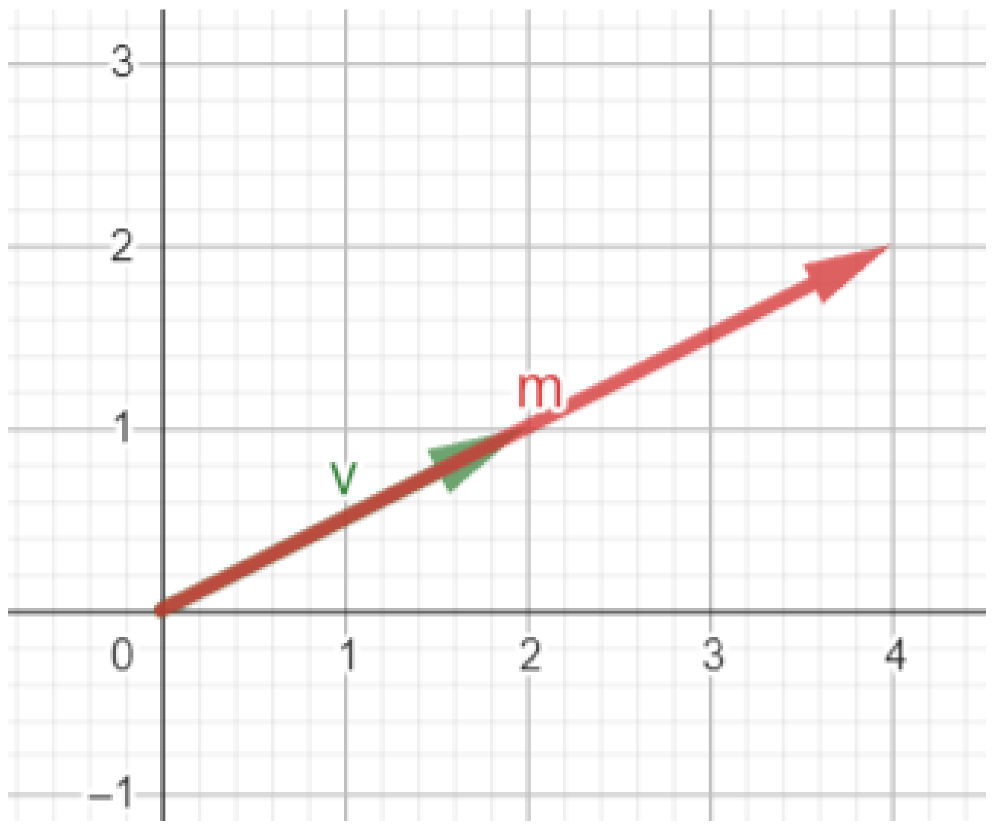
- As a transformation
  - https://www.geogebra.org/calculator/djankxxh

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \quad x = \begin{bmatrix} -2 \\ -2 \end{bmatrix} \quad Ax = \begin{bmatrix} -6 \\ -2 \end{bmatrix} \qquad Ax = b$$

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \quad v = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad Av = \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2\begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad Av = \lambda v$$

- Eigen vectors exist only for square matrices
- They are special
- At most n distinct eigen vectors for nxn matrix

Unit Eigen Vector



$$A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \quad v = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad Av = \lambda v$$

- Infinite eigen vectors along the line !!
- Unit Eigen Vector = $\dfrac{v}{\|v\|} = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$

- Norm of the unit Eigen vector is 1

$$\left(\frac{2}{\sqrt{5}}\right)^2 + \left(\frac{1}{\sqrt{5}}\right)^2 = \frac{4+1}{5} = 1$$

$$\beta_1^2 + \beta_2^2 = 1$$

Matrix Vector multiplication as change of basis

Coordinates of new basis (as represented in standard basis)

Coordinates in standard basis

Coordinates in new basis

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \quad x = \begin{bmatrix} -2 \\ -2 \end{bmatrix} \quad Ax = \begin{bmatrix} -6 \\ -2 \end{bmatrix} \qquad Ax = b$$

Trivia

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \quad v = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad Av = \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad Av = \lambda v$$

- Special case: When lambda is 1, Coordinates in old and new basis are same

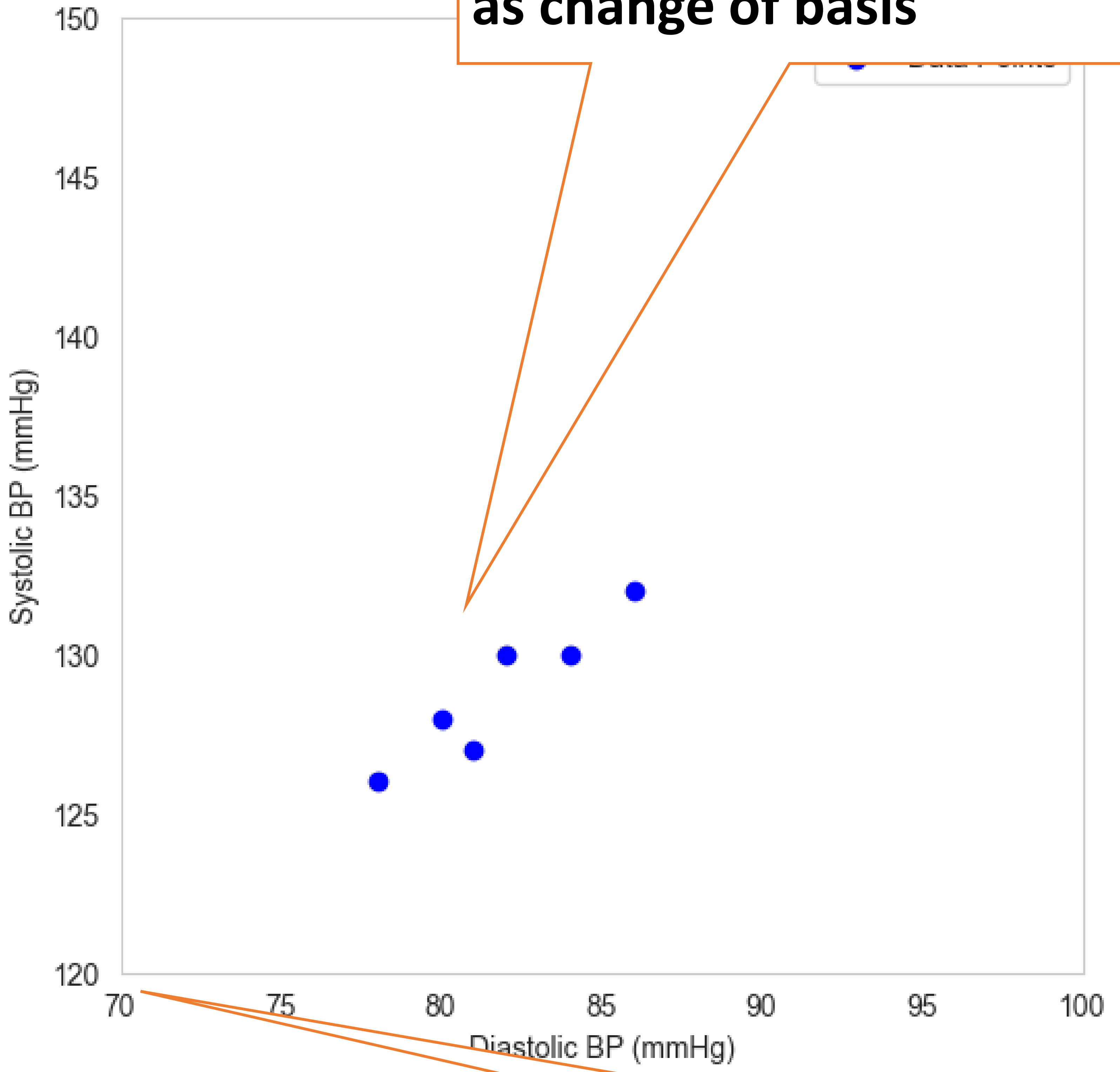
A sample dataset

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

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

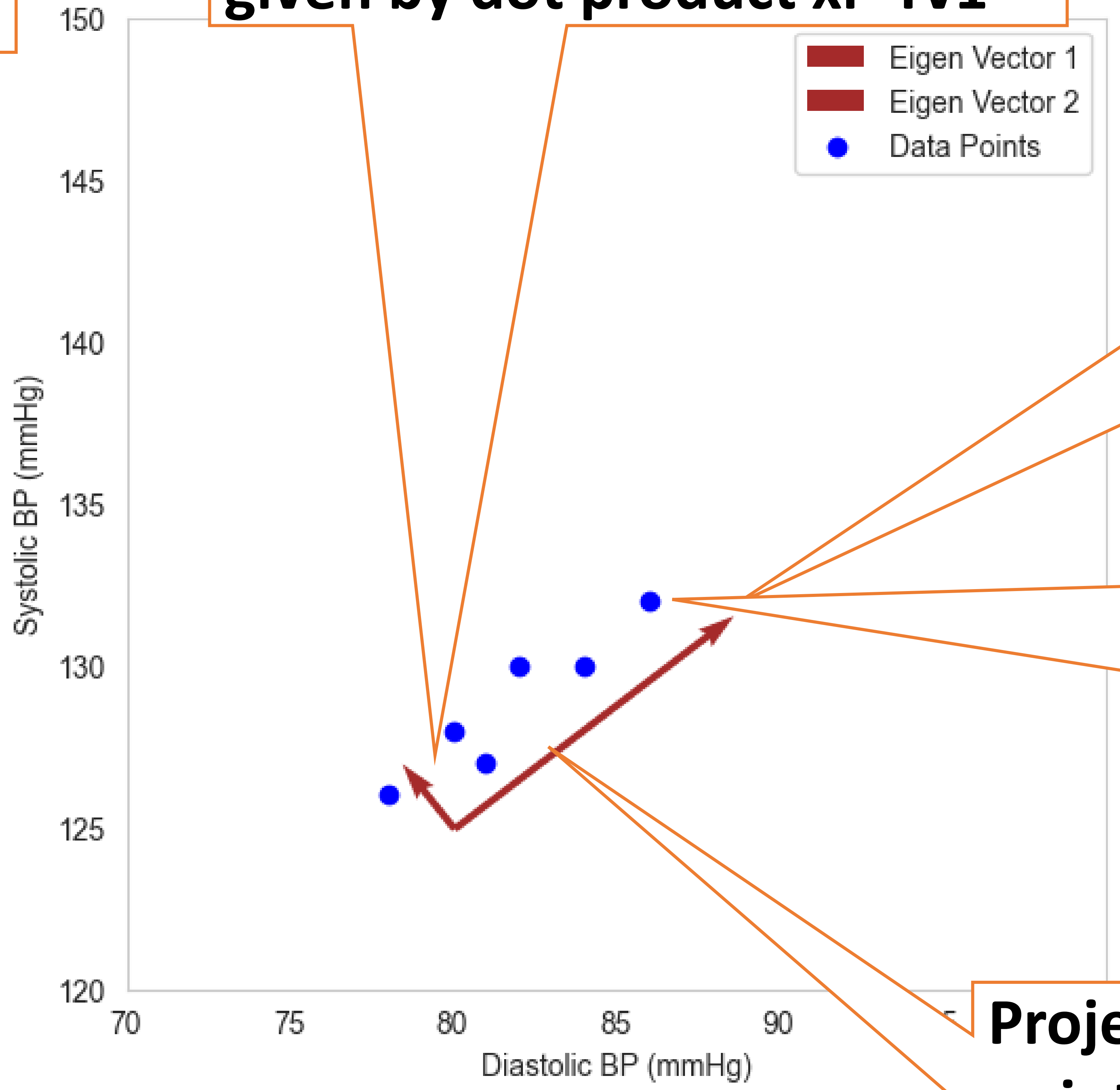
Change of basis (frame of reference)

Centering can also be viewed as change of basis



Standard basis

Projection along new basis is given by dot product xi^Tv1

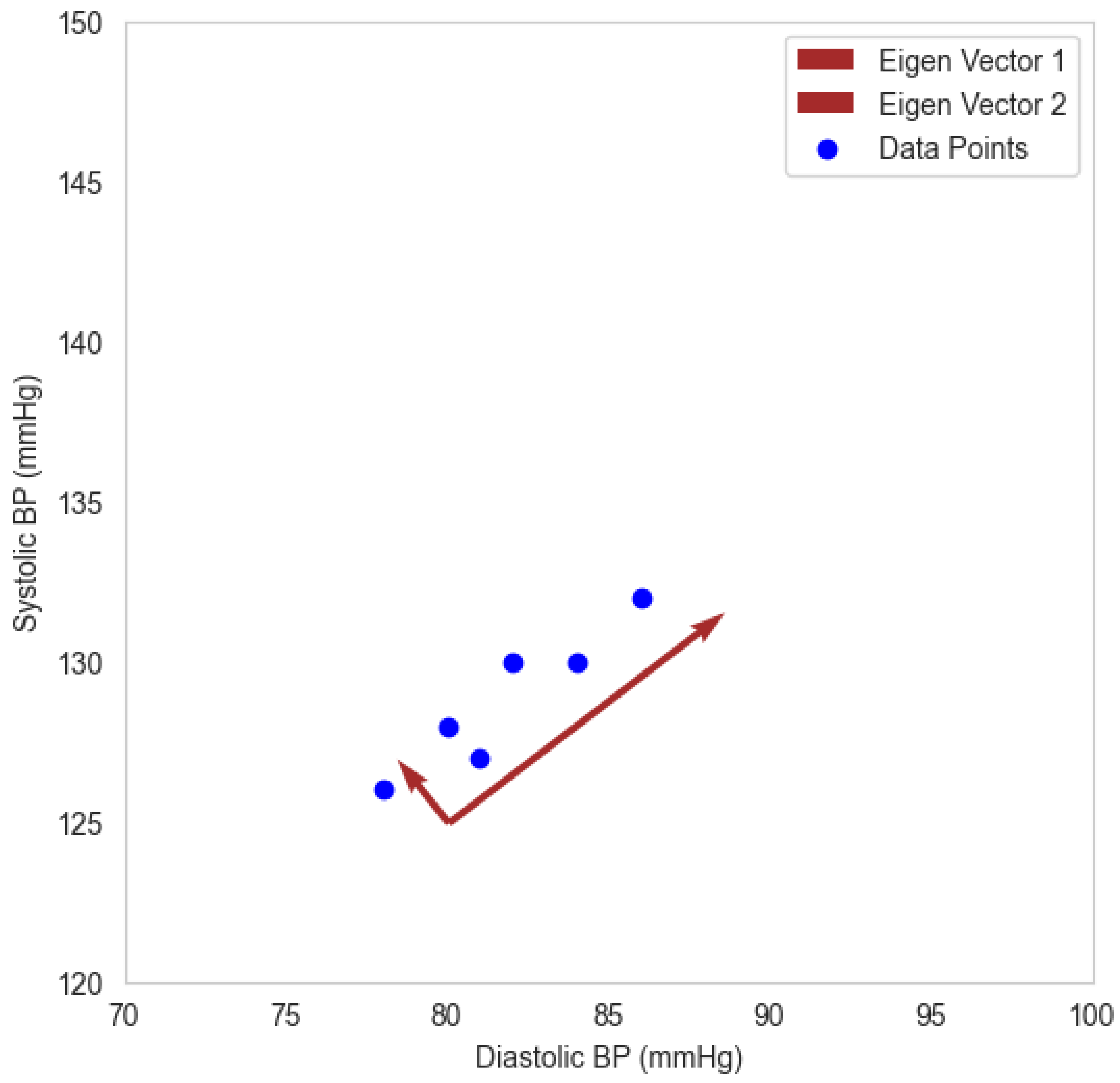


Imagine changing to this new basis

After projection along first vector, very little left

Projection of all points along v is Matrix vector product

Matrix multiplication as change of basis



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

$$v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad v_2 = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

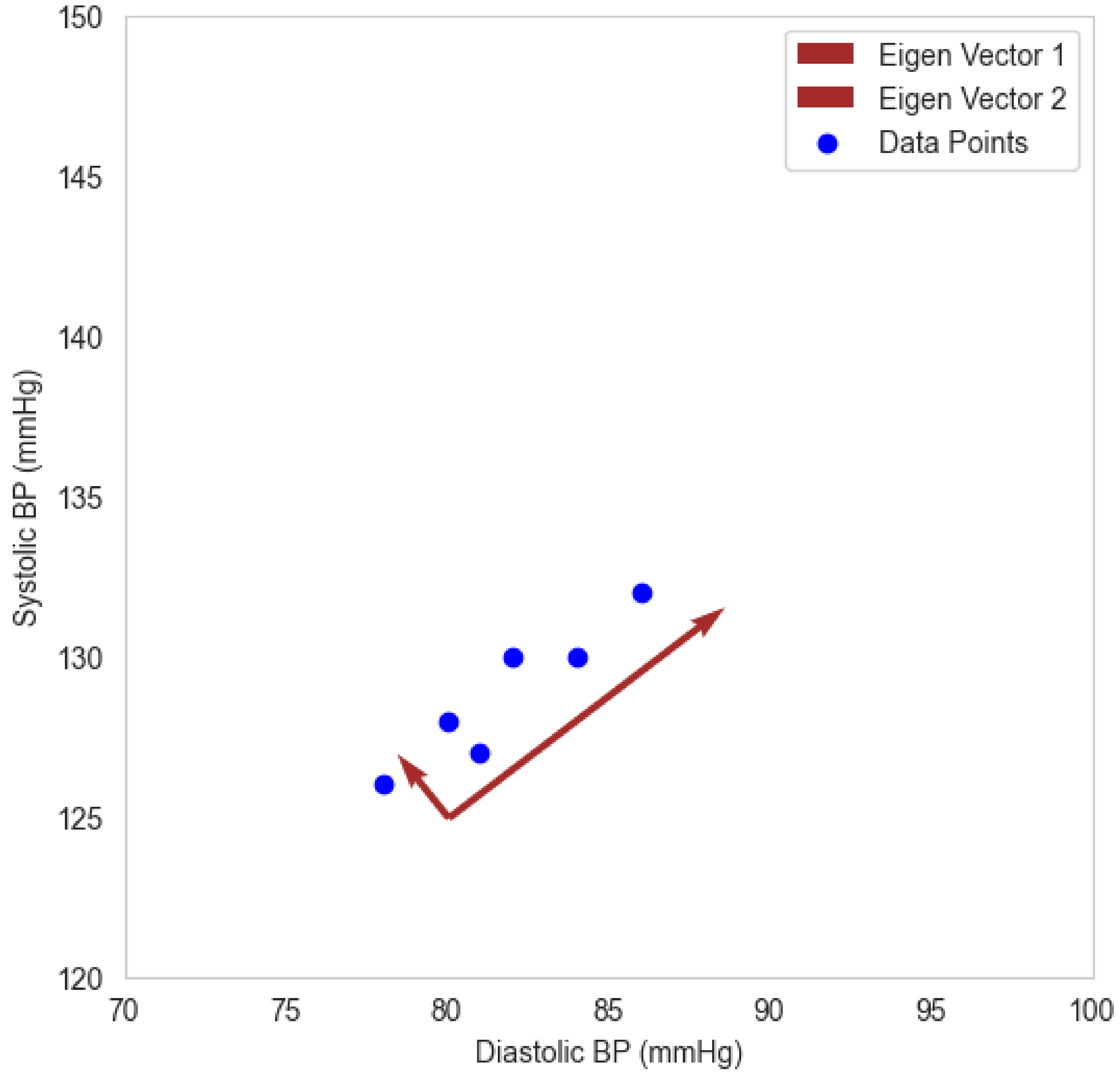
$$\lambda_1 = 12.7 \quad \lambda_2 = 0.38$$

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

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

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

Matrix multiplication as change of basis



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

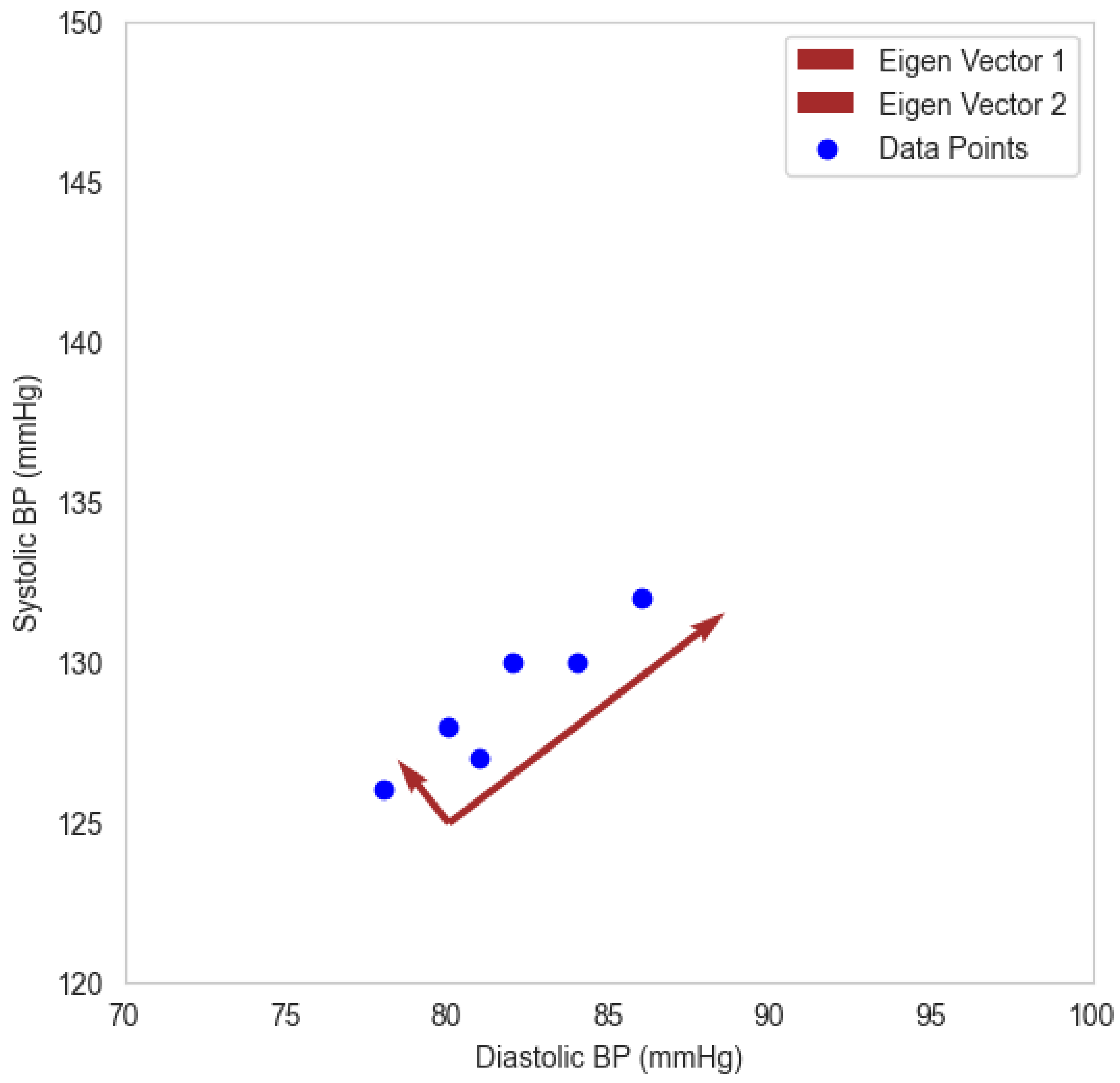
$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

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

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

$$Xv_2 = PC2$$

Matrix multiplication as change of basis



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

$$XV = \begin{bmatrix} \uparrow & \uparrow \\ Xv_1 & Xv_2 \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow \\ PC1 & PC2 \\ \downarrow & \downarrow \end{bmatrix}$$


$X^2 - 4X \leq 0$
$n(B \cap C) = 22$
$n(B) = 68$
$n(C) = 84$
$n(B \cup C) = n(B) + n(C) - n(B \cap C)$
$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$
$\bar{x}_1 = \frac{1+3+3+6+8+9}{6} = 5$
$\bar{x}_2 = 2+4+4+8+12 = 30$
$\bar{x}_3 = 4+7+1+6 = 18$
$\log_b b^x = x$
$\log_a x = \frac{\log_b x}{\log_b a}$
20 →
6 →
He = 4.002602
Na = 22.989769
Ar = 39.948
(0,1)
$M = \frac{0.046765}{3\ 0L}$
$f = \frac{R}{2}$
$Zn_3Sb_2 + 6H_2O \longrightarrow 3Zn(OH)_2 + 2SbH_3$
A
B
C

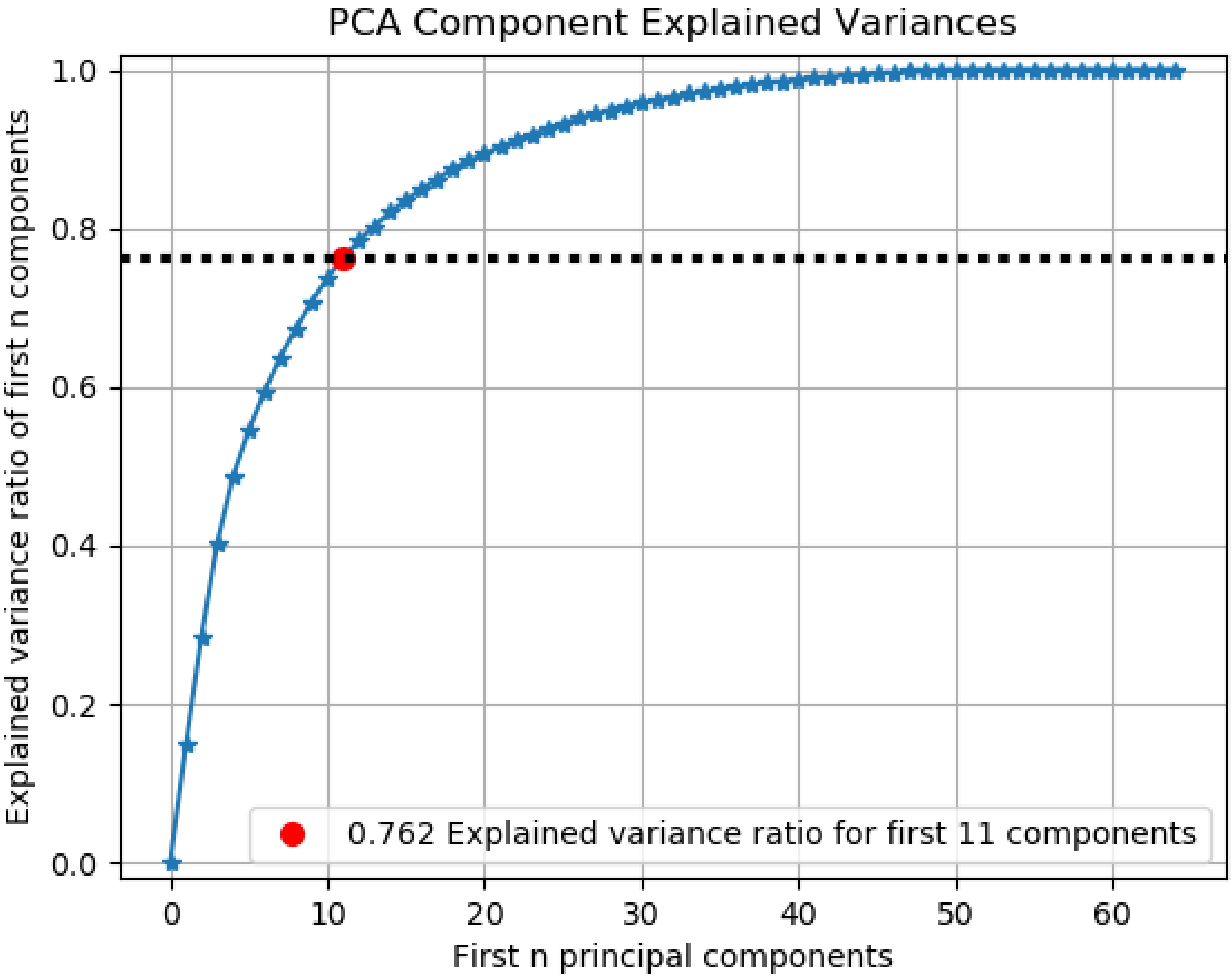
5 minute PCA

PCA steps for dimensionality reduction

- Given a dataset X, perform PCA as follows

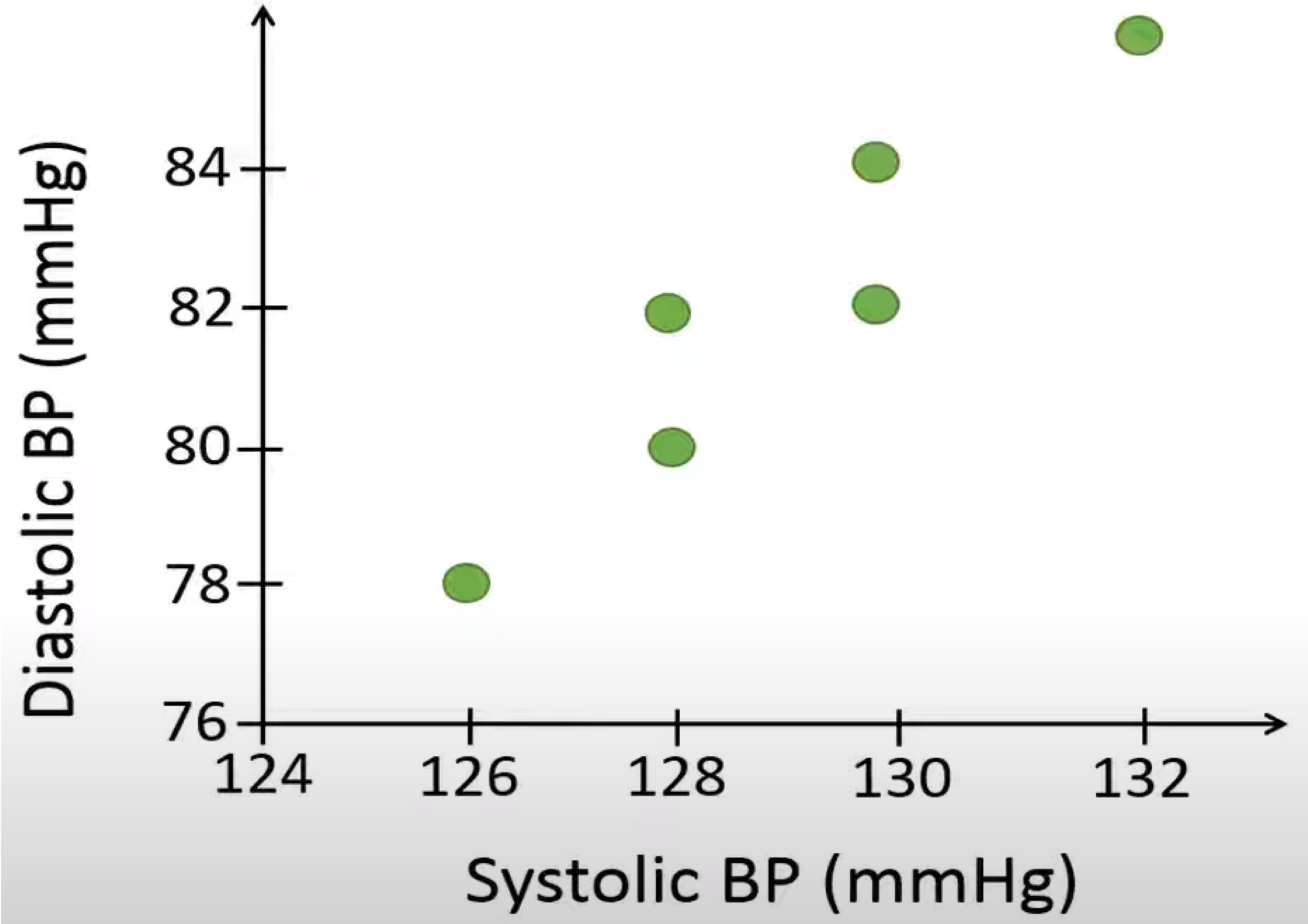
Name	Diastolic BP	Systolic BP
Patient1	78.00	126.00
Patient2	80.00	128.00
Patient3	81.00	127.00
Patient4	82.00	130.00
Patient5	84.00	130.00
Patient6	86.00	132.00

- Calculate covariance matrix of X
- Cov matrix is square
- Calculate Eigen values & Eigen vectors of cov matrix
- Eigen vectors by descending order of eigen value
- X.shape = 6x2, Cov shape = 2 x 2, V shape = 2x2
- Calculate XV, throw away last few columns of result
- How many columns to throw away?


PCA Component Explained Variances
Explained variance ratio of first n components
First n principal components
0.762 Explained variance ratio for first 11 components
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60

Example Data

Systolic BP	Diastolic BP
126	78
128	80
128	82
130	82
130	84
132	86



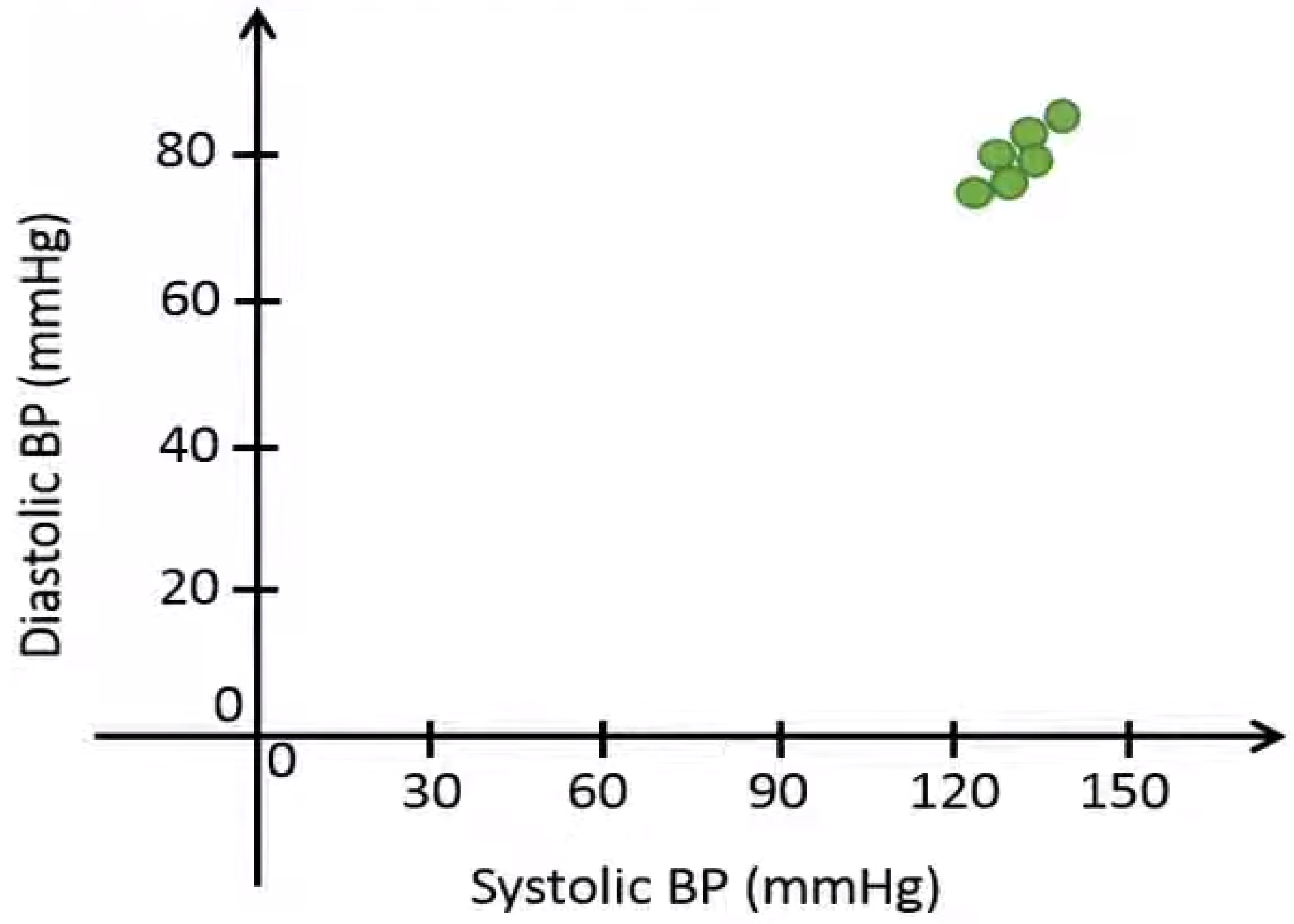
Steps

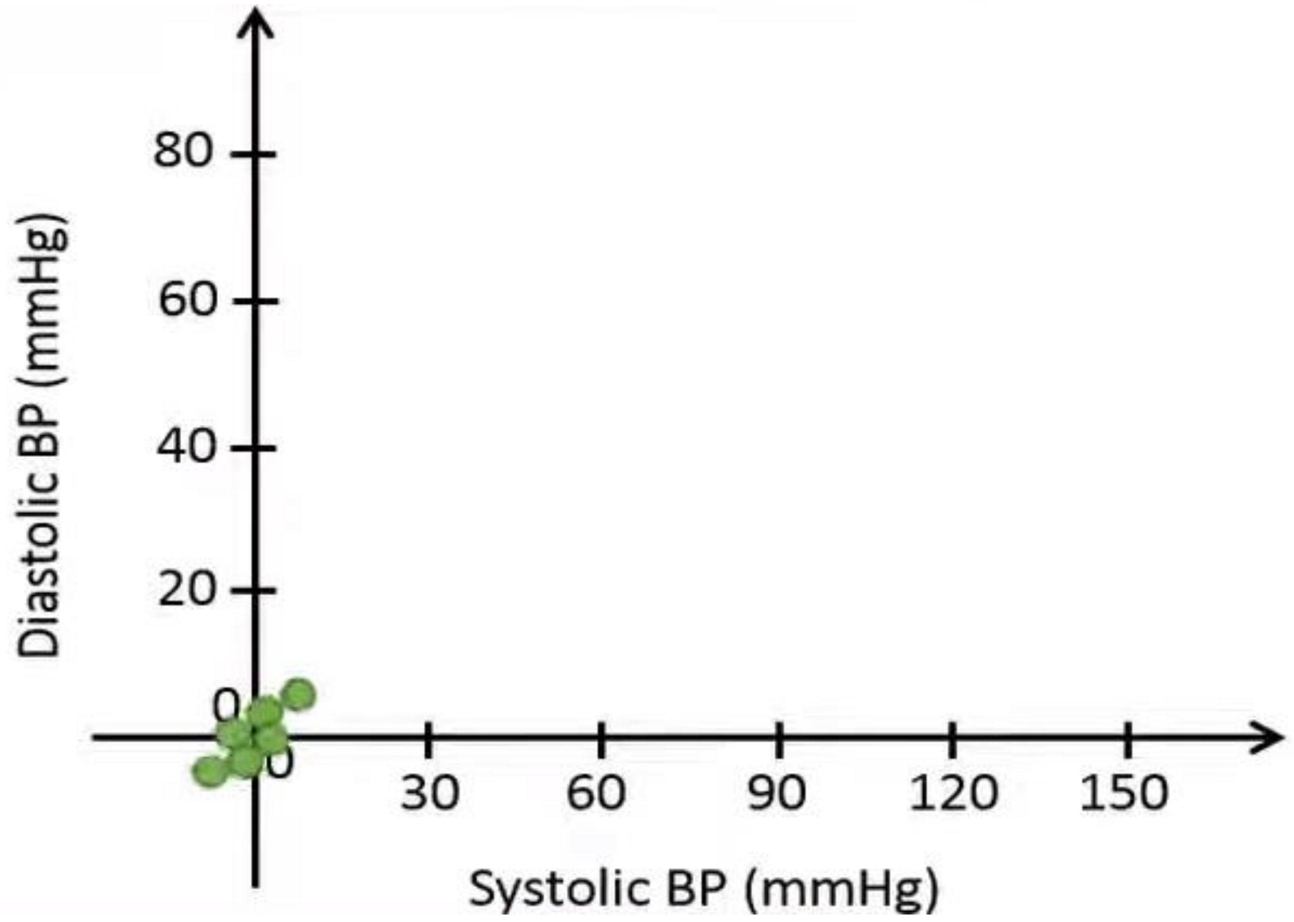
- Center the data (for convenience)
- Calculate covariance matrix
- Calculate eigen values of covariance matrix
- Calculate eigen vectors of covariance matrix
- Order the eigen vectors in descending value of eigen values
- Calculate principal components

Step 1: Center data

Systolic BP	Diastolic BP
126 -129 = -3	78 -82 = -4
128 -129 = -1	80 -82 = -2
128 -129 = -1	82 -82 = 0
130 -129 = 1	82 -82 = 0
130 -129 = 1	84 -82 = 2
132 -129 = 3	86 -82 = 4

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4

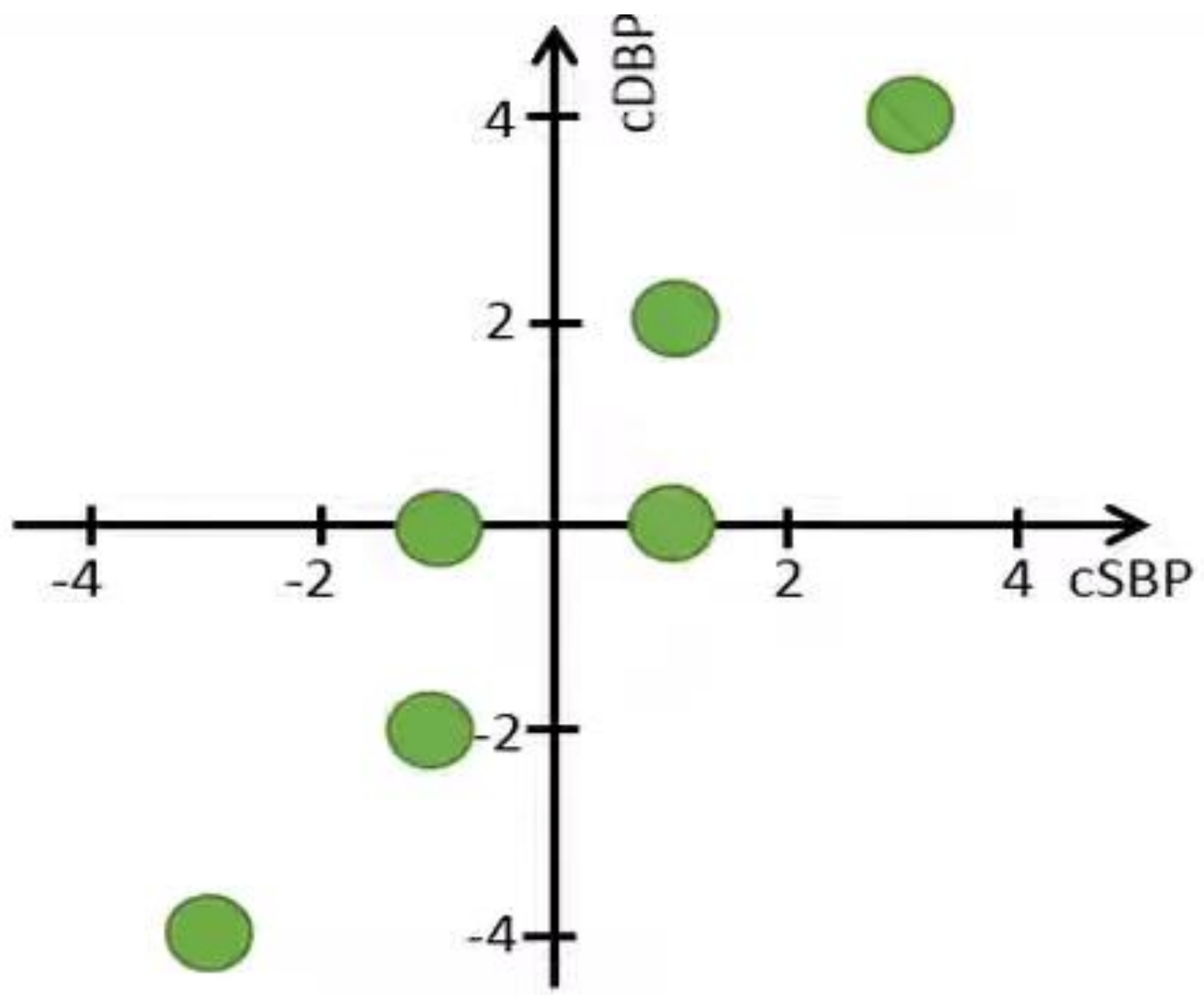




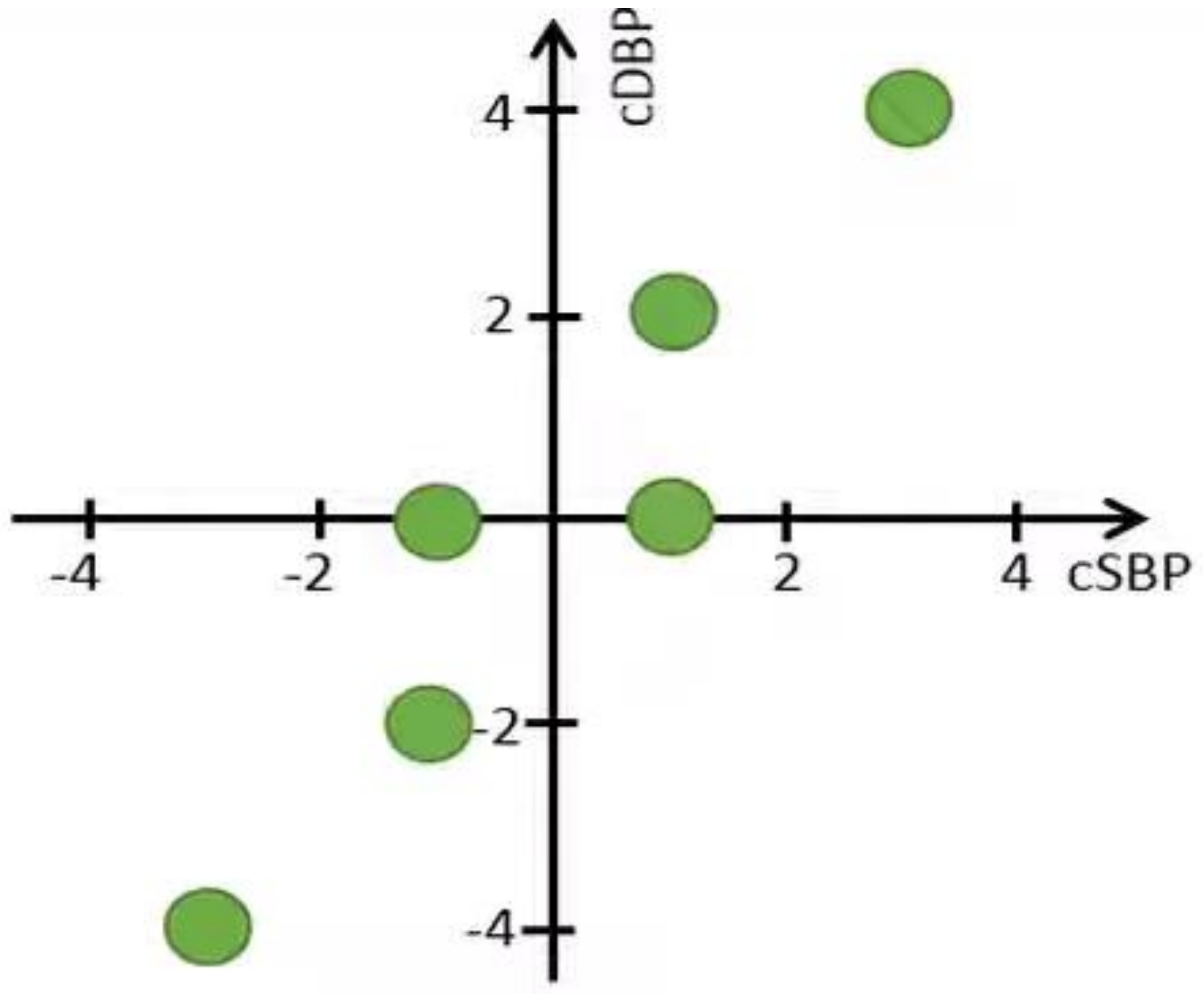
Step 2: Calculate Covariance Matrix

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0



Step 3: Calculate Eigen values of Covariance Matrix



$$\det\left|\boxed{A} - \lambda I\right| = 0$$

$$\det\left|\boxed{\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix}} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right| = 0$$

$$(4.4 - \lambda)(8.0 - \lambda) - 5.6 \cdot 5.6 = 0$$

$$\boxed{3.84 - 12.4\lambda + \lambda^2 = 0}$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\lambda_1 = 0.32 \qquad \lambda_2 = 12.08$$

Step 4: Calculate Eigen vectors of Covariance Matrix

$$\lambda_1 = 0.32 \qquad \lambda_2 = 12.08$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \boxed{0.32} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

$$A \cdot v = \lambda \cdot v$$

$$\boxed{\begin{bmatrix} 4.4 & 5.6 \\ 5.6 & 8.0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}} = 12.08 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

$$4.4x + 5.6y = 12.08x$$

$$5.6x + 8.0y = 12.08y$$

$$5.6y = 7.68x$$

$$5.6x = 4.08y$$

$$\boxed{\begin{aligned} y &= 1.37x \\ 1.37x &= y \end{aligned}}$$

$$v_2 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \quad \lambda_2 = 12.08$$

$$v_1 = \begin{bmatrix} -0.81 \\ 0.59 \end{bmatrix} \quad \lambda_1 = 0.32$$

Step 5: Reorder eigen vectors

$$\lambda_1 = 0.32 \qquad \lambda_2 = 12.08$$

	SBP	DBP
SBP	4.4	5.6
DBP	5.6	8.0

$$v_1 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \quad \lambda_1 = 12.08$$

$$v_2 = \begin{bmatrix} -0.81 \\ 0.59 \end{bmatrix} \quad \lambda_2 = 0.32$$

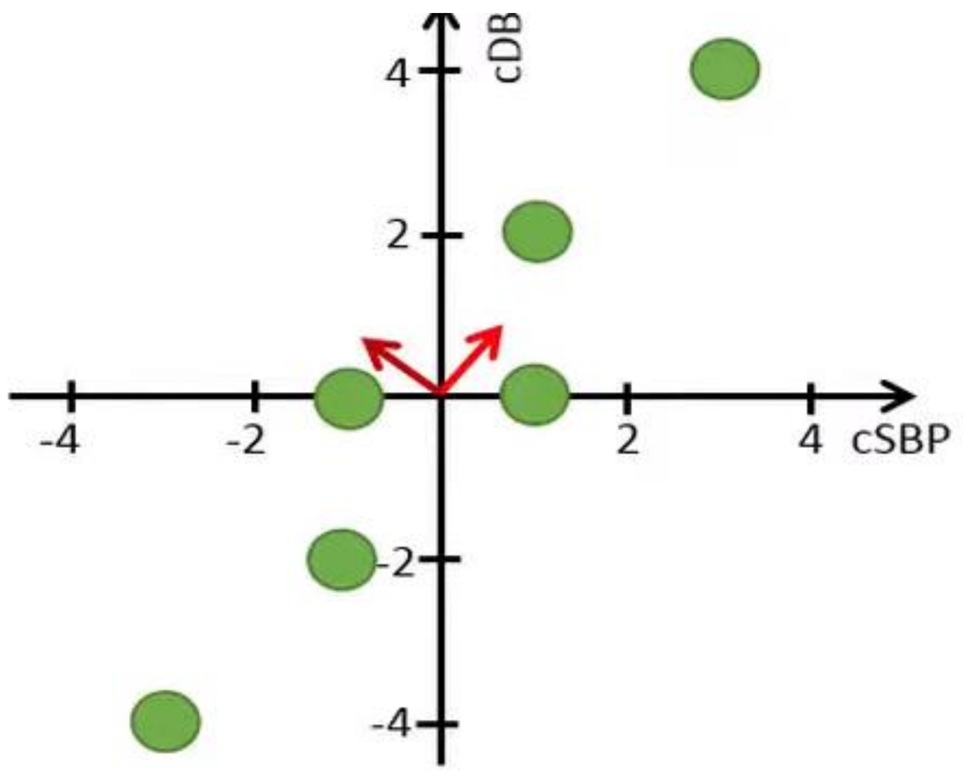
$$v_2 = \begin{bmatrix} 0.59 \\ 0.81 \end{bmatrix} \quad \lambda_2 = 12.08$$

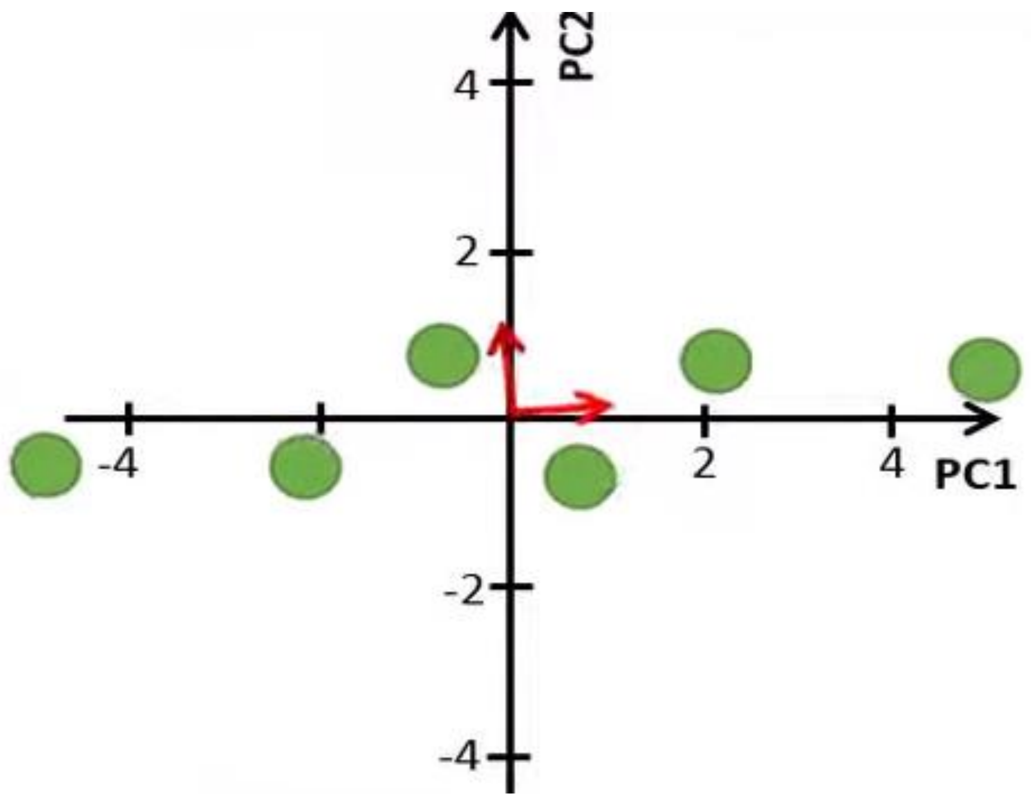
$$v_1 = \begin{bmatrix} -0.81 \\ 0.59 \end{bmatrix} \quad \lambda_1 = 0.32$$

$$V = \begin{bmatrix} \boxed{\begin{matrix} 0.59 \\ 0.81 \end{matrix}} & \begin{matrix} -0.81 \\ 0.59 \end{matrix} \end{bmatrix}$$

Step 6: Calculate Principal Components

Centered SBP	Centered DBP
-3	-4
-1	-2
-1	0
1	0
1	2
3	4



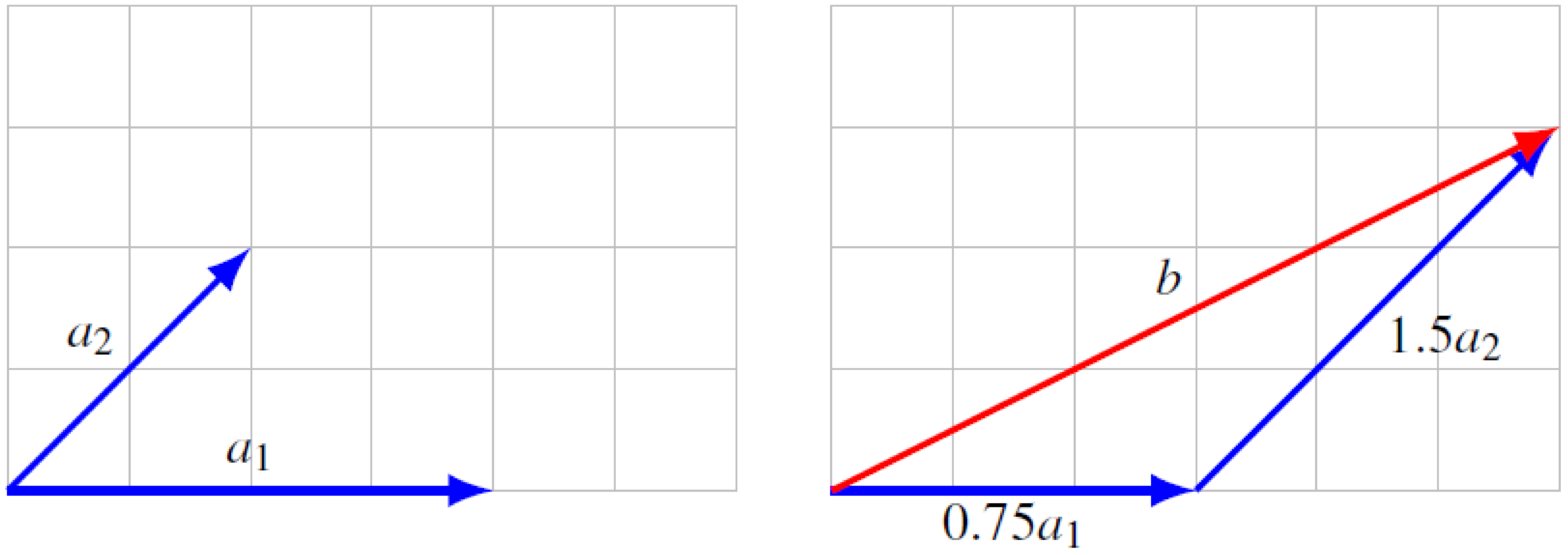


$$X = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad V = \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix}$$

$$XV = \begin{bmatrix} -3 & -4 \\ -1 & -2 \\ -1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 0.59 & -0.81 \\ 0.81 & 0.59 \end{bmatrix} = \begin{bmatrix} -5.0 & 0.1 \\ -2.2 & -0.4 \\ -0.6 & 0.8 \\ 0.6 & -0.8 \\ 2.2 & 0.4 \\ 5.0 & -0.1 \end{bmatrix}$$

Linear Combination

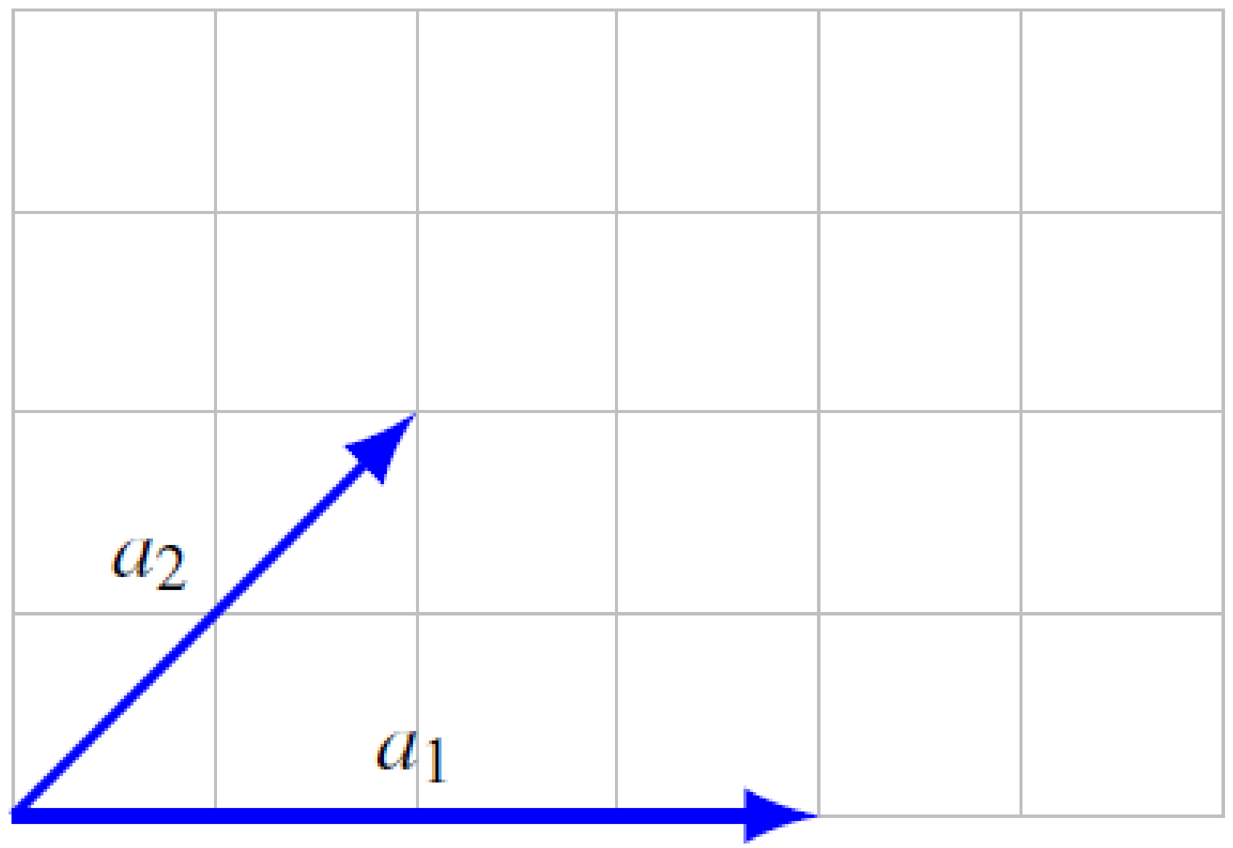
Linear Combinations

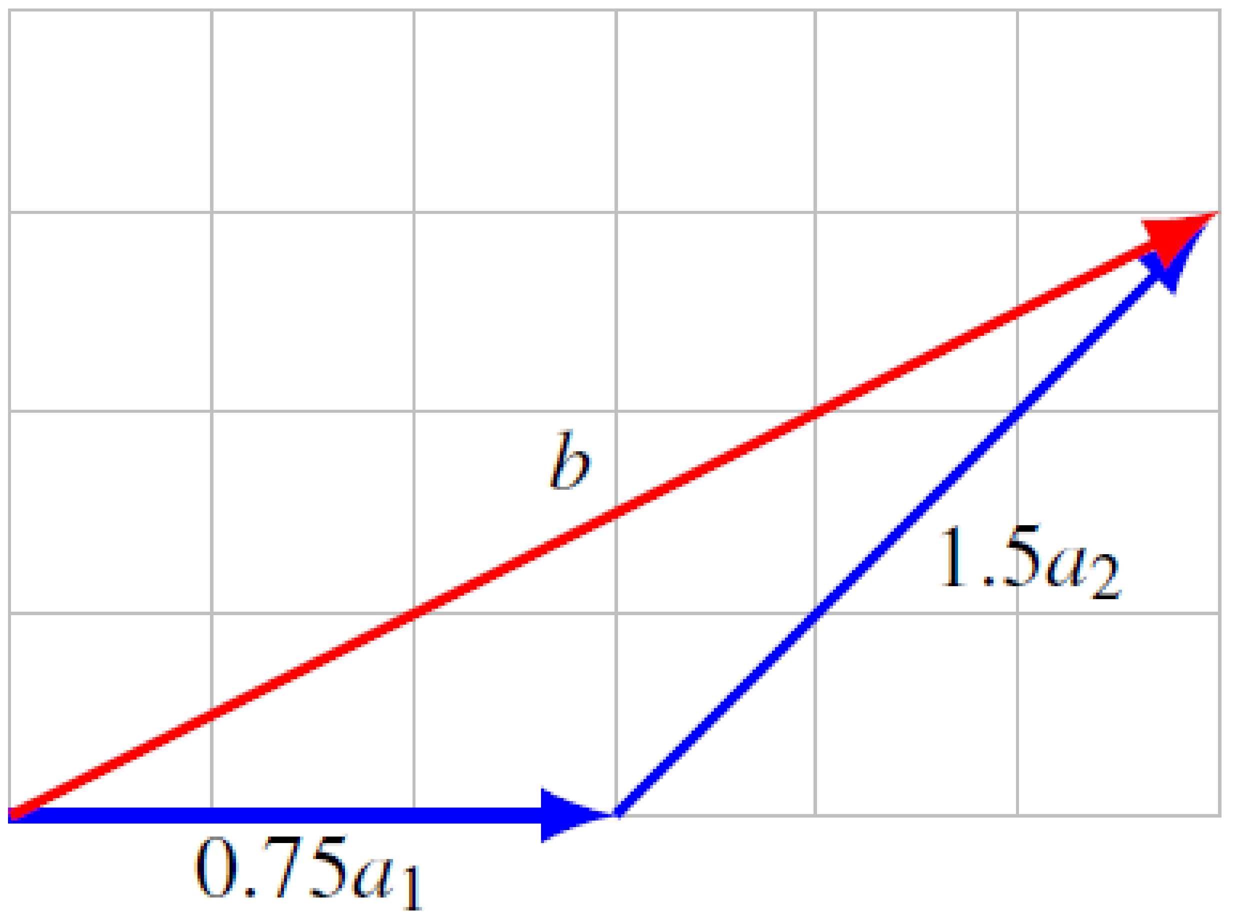


- Scale and add vectors

Linear Combinations

- Definition: $\beta_1\mathbf{x_1} + \beta_2\mathbf{x_2} + \dots + \beta_n\mathbf{x_n}$
  - $\beta_1, \beta_2, \dots, \beta_n$ are scalars
  - $\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_n}$ 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

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

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

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

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 intuition

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.

PCA Goals

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

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

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

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

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

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

$$PCi = ...$$

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

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

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

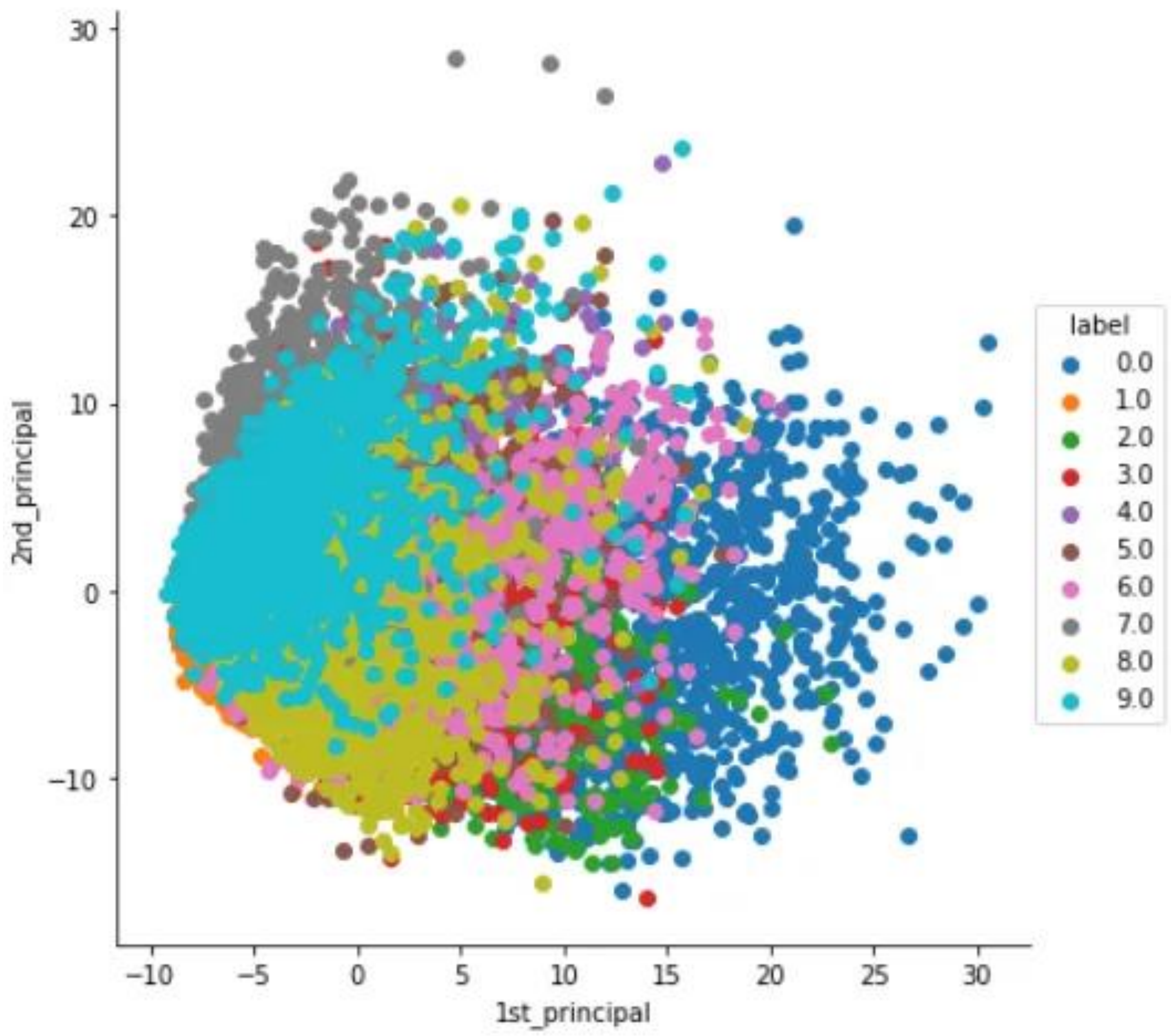

$X^2 - 4X \le 0$
$n(B \cap C) = 22$
$n(B) = 68$
$n(C) = 84$
$n(B \cup C) = n(B) + n(C) - n(B \cap C)$
$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$
$\bar{x}_1 = \frac{1+3+3+6+8+9}{6} = 5$
$\bar{x}_2 = 2+4+4+8+12 = 30$
$\bar{x}_3 = 4+7+1+6 = 18$
$\log_b b^x = x$
$\log_a x = \frac{\log_b x}{\log_b a}$
$20 \longrightarrow$
$6 \longrightarrow$
He = 4.002602
Na = 22.989769
Ar = 39.948
$M = \frac{0.046765}{3\,0L}$
(0,1)
$f = \frac{R}{2}$
$Zn_3Sb_2 + 6H_2O \longrightarrow 3Zn(OH)_2 + 2SbH_3$

Beyond PCA

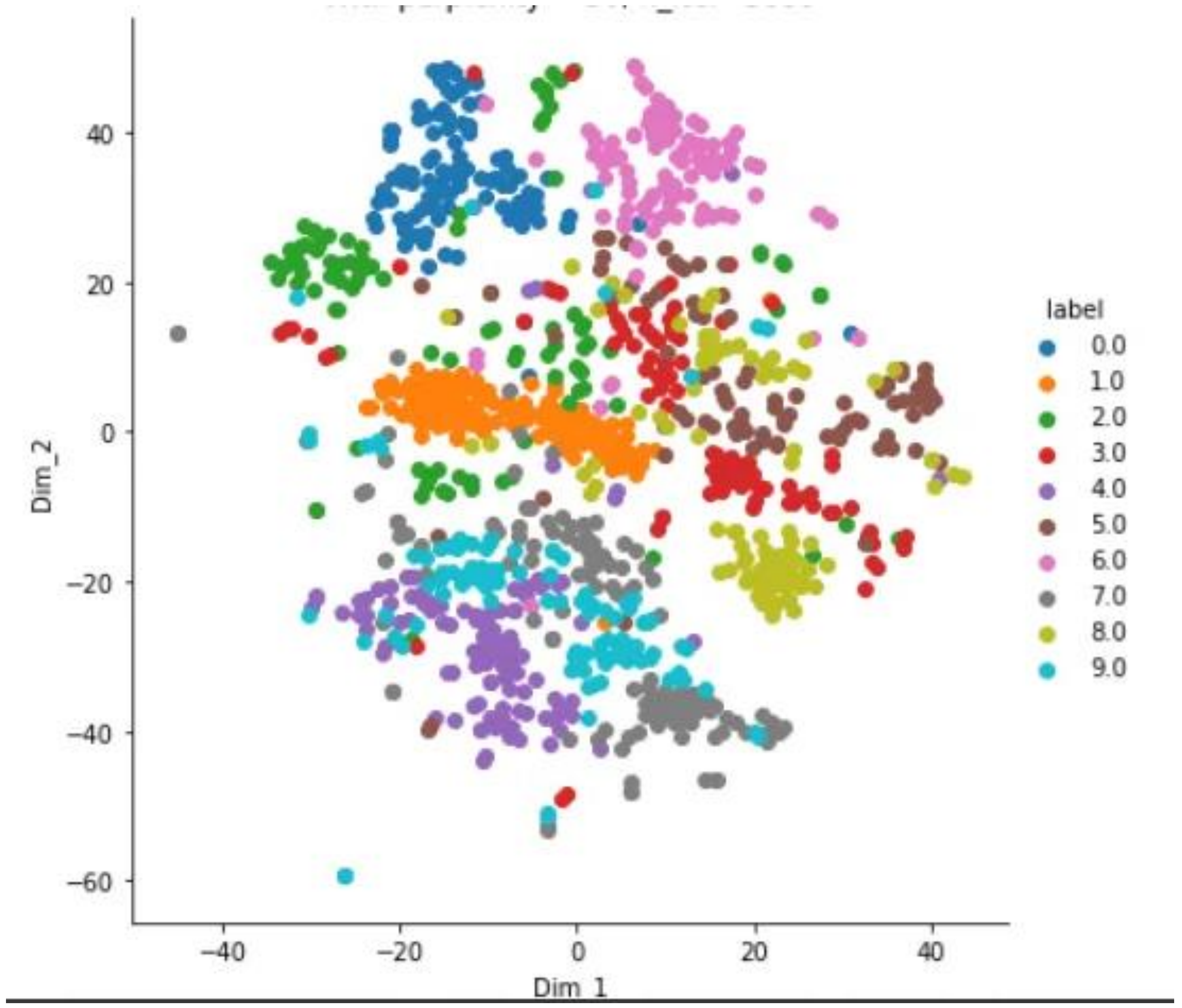
Non linear dimensionality reduction methods

- Optional
- t-SNE
  - https://www.youtube.com/watch?v=MnRskV3NY1k
- Precursors to t-SNE: IsoMap, MDS
  - https://towardsdatascience.com/manifold-learning-t-sne-lle-isomap-made-easy-42cfd61f5183
- UMAP

With PCA



With t-SNE

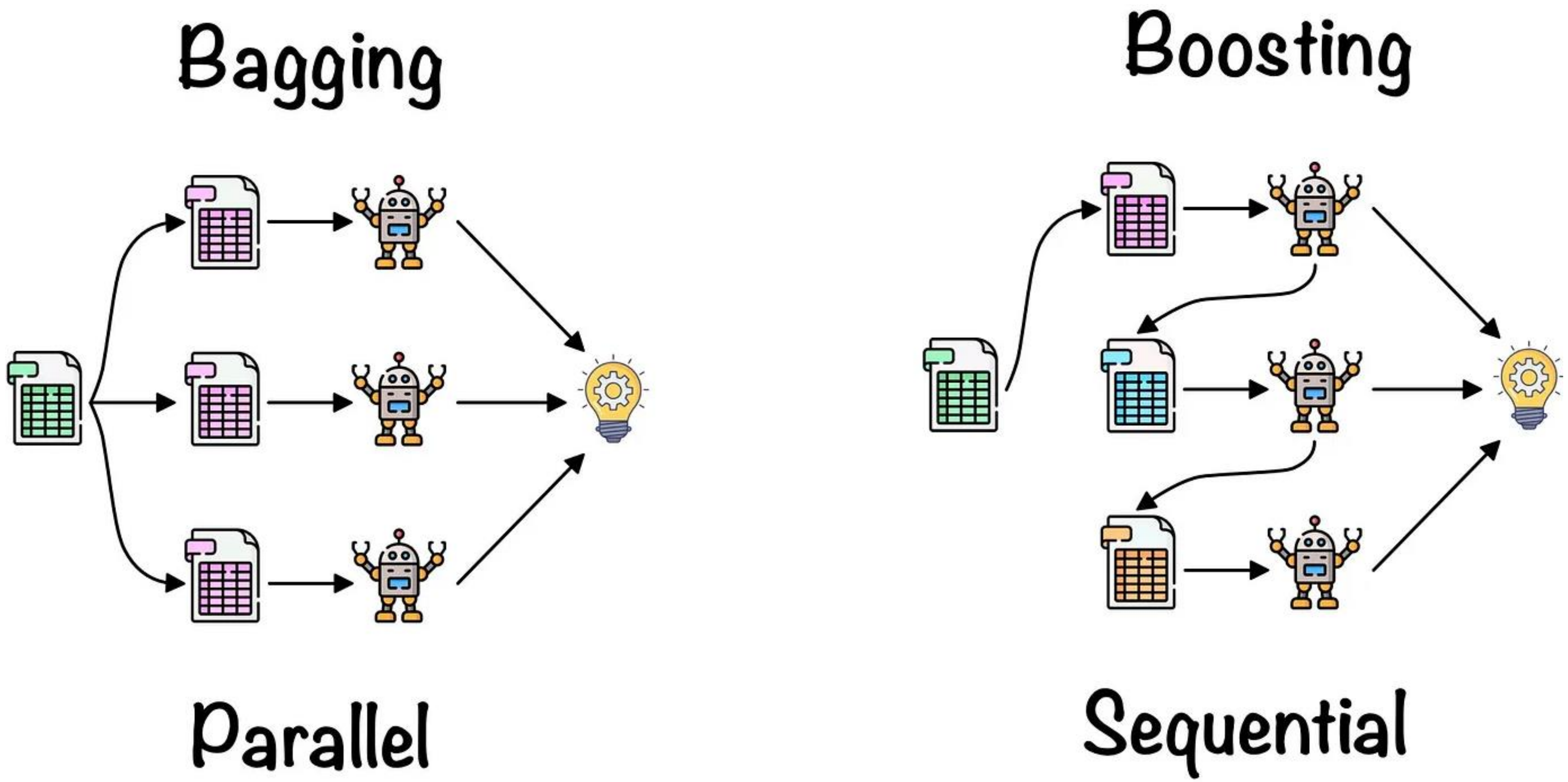


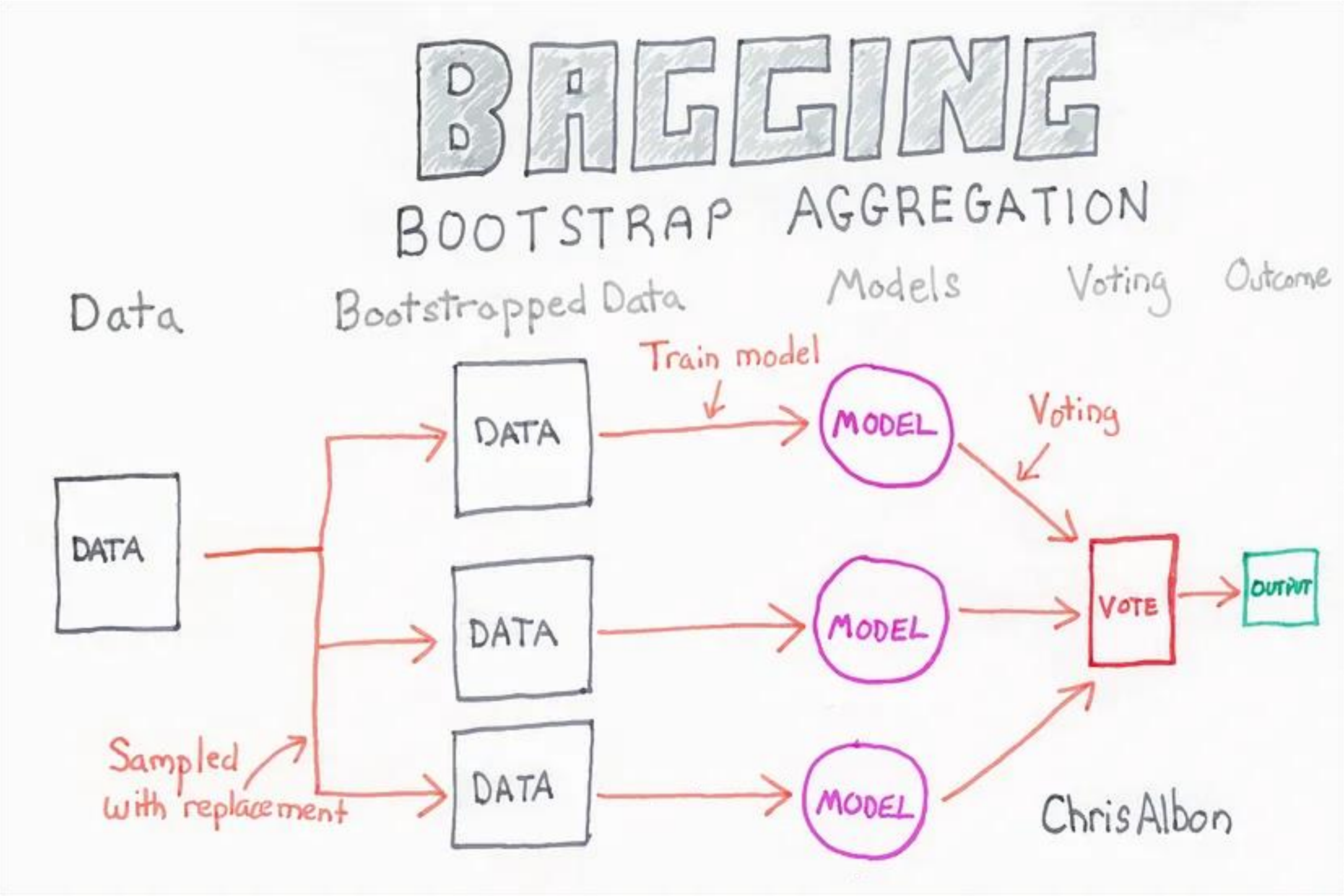

X²– 4X ≤ 0
n(B∩C) = 22
n(B) = 68
n(C) = 84
n(B∪C) = n(B) + n(C)-n(B∩C)
x̄₁ = 1+3+3+6+8+9 / 6 = 5
x̄₂ = 2+4+4+8+12 = 30
x̄₃ = 4+7+1+6 = 18
log_b b^x = x
log_a x = log_b x / log_b a
20 →
6 →
He = 4.002602
Na = 22.989769
Ar = 39.948
M = 0.046765 / 3 0L
(0,1)

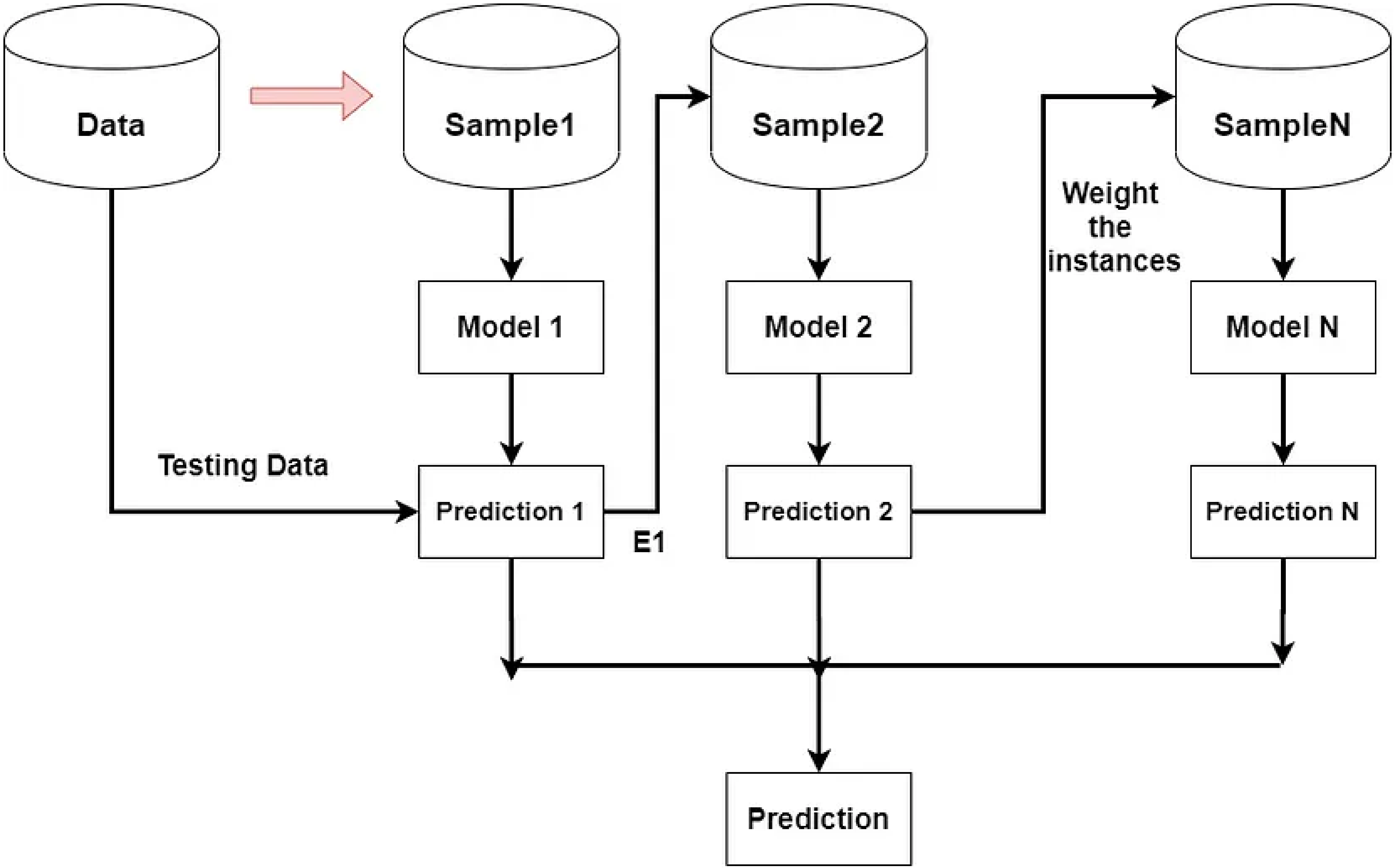
Boosting

Ensemble Learning

- Multiple ML model used together for prediction




BAGGING
BOOTSTRAP AGGREGATION
Data
Bootstrapped Data
Models
Voting
Outcome
Train model
DATA
DATA
DATA
DATA
MODEL
MODEL
MODEL
Voting
VOTE
OUTPUT
Sampled with replacement
Chris Albon


Data
Sample1
Sample2
SampleN
Weight the instances
Model 1
Model 2
Model N
Testing Data
Prediction 1
E1
Prediction 2
Prediction N
Prediction

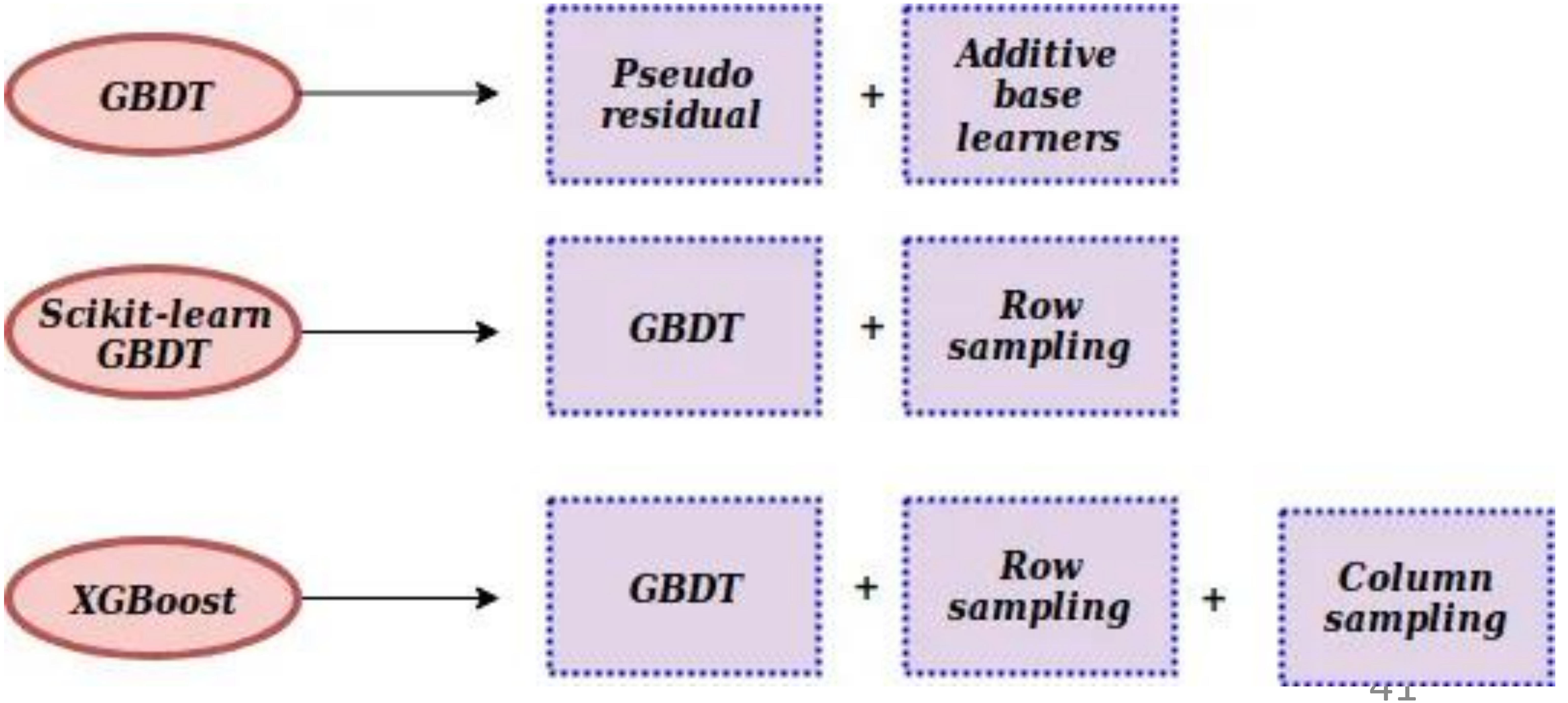
Boosting and Bias/Variance

- Boosting reduces Bias
- Hence low variance models are used as individual learner
  - E.g. Decision Tree with very less depth

- Combine high bias low variance models (weak learners) to make low bias low variance models

Boosting

- AdaBoost
- GradientBoost
- XGBoost



Helpful videos on boosting (Optional)

- AdaBoost
  - https://www.youtube.com/watch?v=LsK-xG1cLYA
- Gradient Boost (4 parts)
  - https://www.youtube.com/watch?v=3CC4N4z3GJc
- XGBoost (4 parts)
  - https://www.youtube.com/watch?v=OtD8wVaFm6E


QUESTIONS