

Here we will find the value of max  $u_1$ , using **Eigen values and Eigen vectors**.

Wkt, co-variance matrix of  $X = S$ , where  $S$  is symmetric matrix.

And for  $d$  no of features there will be  **$d$  eigen-values** and  **$d$  eigen-vector**.

Eigen values are represented by **lambda**, and Eigen vectors are represented by **v**.

Solution to our Optimization problems:  $\lambda_1, V_1$

$X = \frac{1}{2} \begin{bmatrix} 1 & 2 & 3 & \dots & d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n & & & & \end{bmatrix}$   $n \times d$

Covariance matrix of  $X = S$

$S_{d \times d} = X^T X$   $d \times n \quad n \times d$

Sq. Symm. matrix

eigen-values ( $\lambda_1, \lambda_2, \dots, \lambda_d$ )

eigen-vector ( $V_1, V_2, \dots, V_d$ )

Lambda and V are eigen values and vectors of  $S$  respectively, only if:

$$\lambda_1 * V_1 = S * V_1$$

And  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \dots \geq \lambda_d$

$V_1 \geq V_2 \geq V_3 \geq \dots \geq V_d$ , here each eigen vector is linked to each eigen value

$S_{d \times d}$

maximal eigen-value  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \dots \geq \lambda_d$

eigen-values of  $(S) = \lambda_1, \lambda_2, \lambda_3, \lambda_4 \dots \lambda_d$

eigen vectors of  $(S) = V_1, V_2, V_3, V_4 \dots V_d$

def:  $\lambda_1 V_1 = S V_1$

$\lambda_1$ : scalar

$V_1$ :  $d \times 1$  vector

$\lambda_1 V_1 = S V_1$

$\lambda_1$ : eigen value of  $S$

$V_1$ : eigen vec to  $S$  corr. to  $\lambda_1$

Each Eigen vector is perpendicular to another, that means their dot product is 0, since  $\cos 90^\circ$  is 0.

And  $\mathbf{u}_1 = \mathbf{V}_1$ , where

$\mathbf{U}_1$  is max-variance direction,

$\mathbf{V}_1$  is eigen vector of  $\mathbf{S}$ , corresponding to largest eigen value  $\lambda_1$

Handwritten notes on a blackboard:

- At the top, eigenvalues are listed in descending order:  $\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_d$ . Below them, corresponding eigenvectors are listed:  $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \dots, \mathbf{V}_d$ . Arrows point from each eigenvalue to its corresponding eigenvector.
- To the right, the matrix  $\mathbf{S}_{d \times d}$  is written.
- In the middle, a box contains  $\mathbf{V}_i \perp \mathbf{V}_j$ , followed by the equation  $\mathbf{V}_i^T \mathbf{V}_j = 0 = \mathbf{V}_i \cdot \mathbf{V}_j = 0$ .
- Below the box,  $\mathbf{u}_1 = \mathbf{V}_1$  is written, with a note that  $\mathbf{u}_1$  is the "eigen-vector of  $\mathbf{S} (= \mathbf{X}^T \mathbf{X})$  corr. to largest eigen-value ( $= \lambda_1$ )".
- An arrow points from the text "max-variance direction" to the circled  $\mathbf{u}_1$ .

### Steps to find $\mathbf{u}_1$ using eigen vectors and values

1. Do Column standardization of  $\mathbf{X}$ .
2. Find co-variance matrix  $\mathbf{S} = \mathbf{X}^T \mathbf{X}$ .
3. Find eigen values ( $\lambda_i$ ) and eigen vectors ( $\mathbf{V}_i$ ) of  $\mathbf{S}$ .
4. Then  $\mathbf{u}_1$  will be equal to  $\mathbf{V}_1$ .

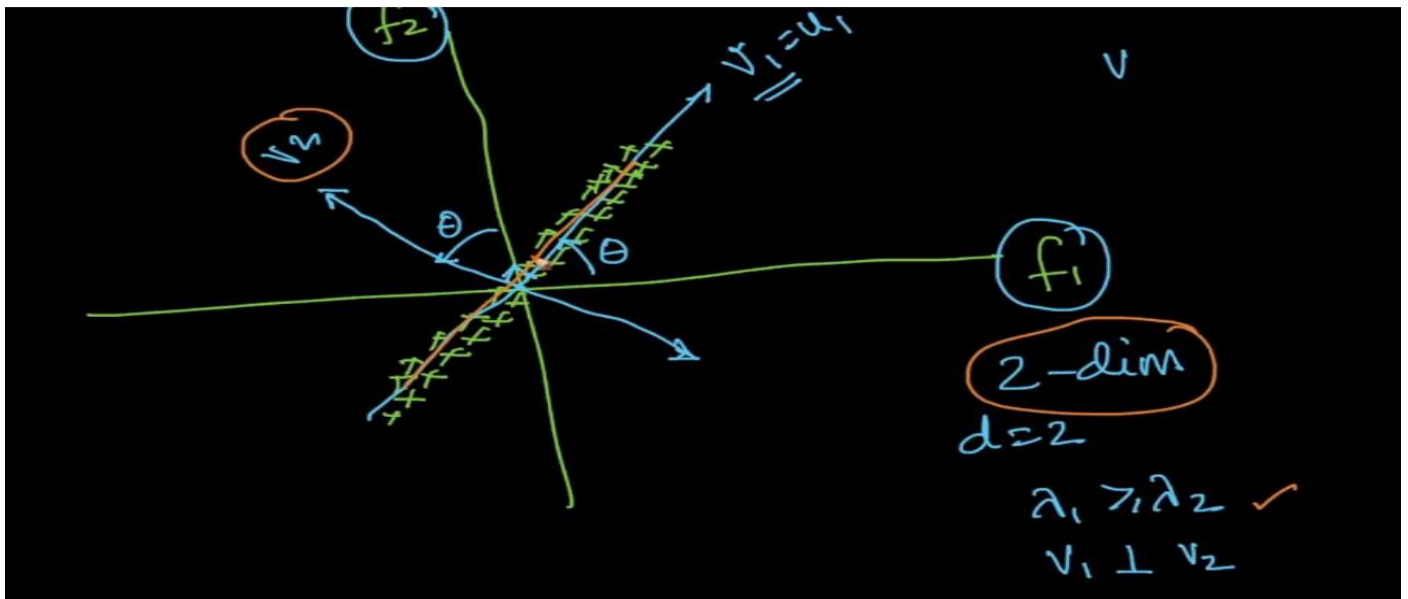
Handwritten notes on a blackboard summarizing the steps:

- On the left,  $\mathbf{X} = \begin{bmatrix} \checkmark \end{bmatrix}_{n \times d}$  is written.
- A large curly bracket groups four numbered steps:
  - ① Col. std of  $\mathbf{X}$  is done
  - ②  $\mathbf{S}_{d \times d} = \mathbf{X}^T \mathbf{X}$
  - ③ eigen values & vectors of  $\mathbf{S}$ . Below this,  $\lambda_1 > \lambda_2 > \dots > \lambda_d$  and  $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_d$  are listed.
  - ④  $\mathbf{u}_1 = \mathbf{V}_1$  (why?)
- At the bottom left,  $\lambda_i$ 's &  $\mathbf{V}_i$ 's are written with a small circle around them.

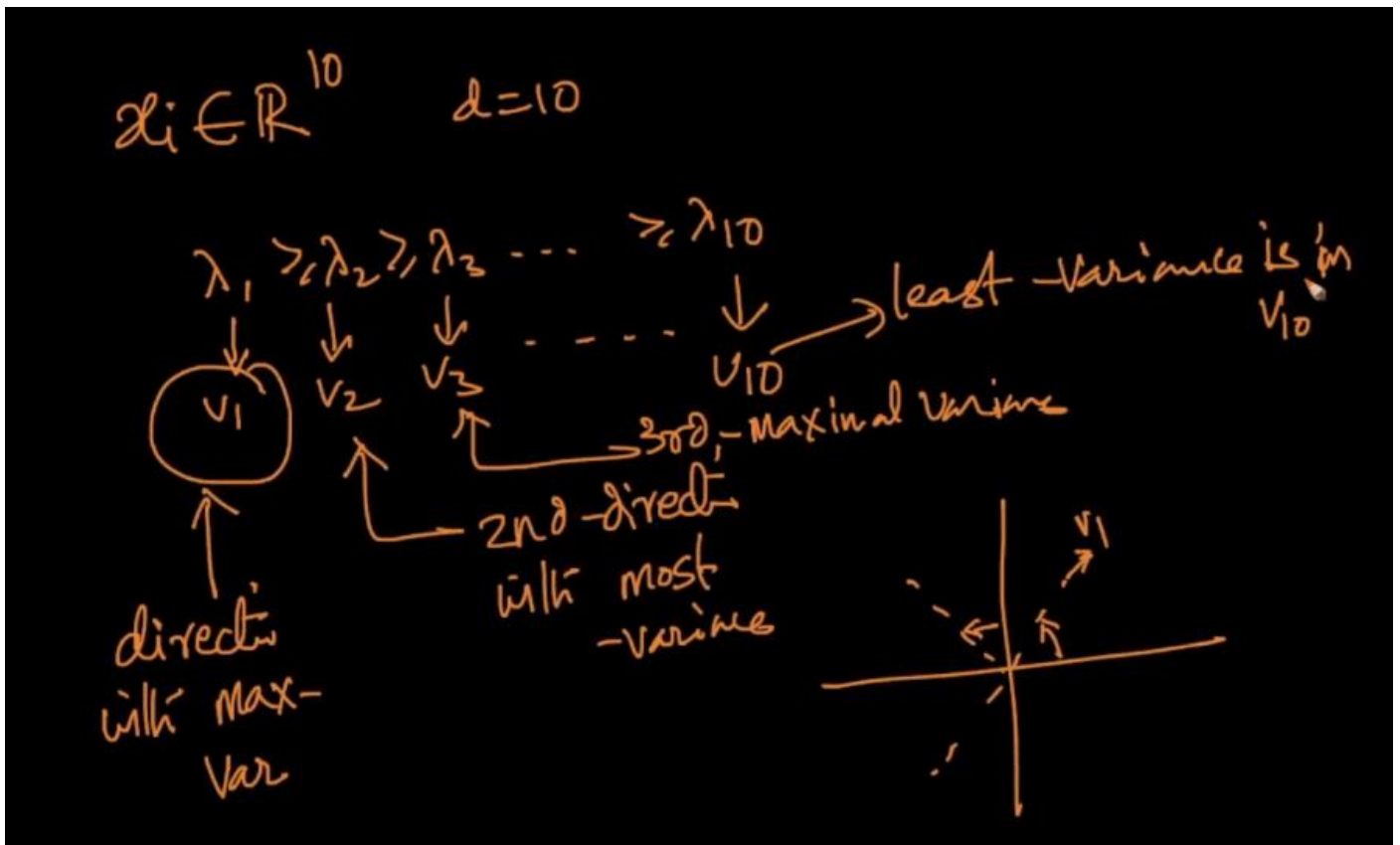
Why  $u_1 = v_1$ ?

Below figure shows  $V_1$  for 2-dimensions, and it's perfectly the containing all the  $x$ 's.

Here  $v_2$  is perpendicular to  $v_1$ .



Now above image shows only for 2-dimensions, but for more dimensions let's say 10, we will have 10 eigen values and 10 eigen vectors, where  $V_1$  will be the direction with max-variance i.e. ( $U_1$ ),  $V_2$  will be 2<sup>nd</sup> maximal variance direction,  $V_3$  will be 3<sup>rd</sup> maximal variance direction and so on.

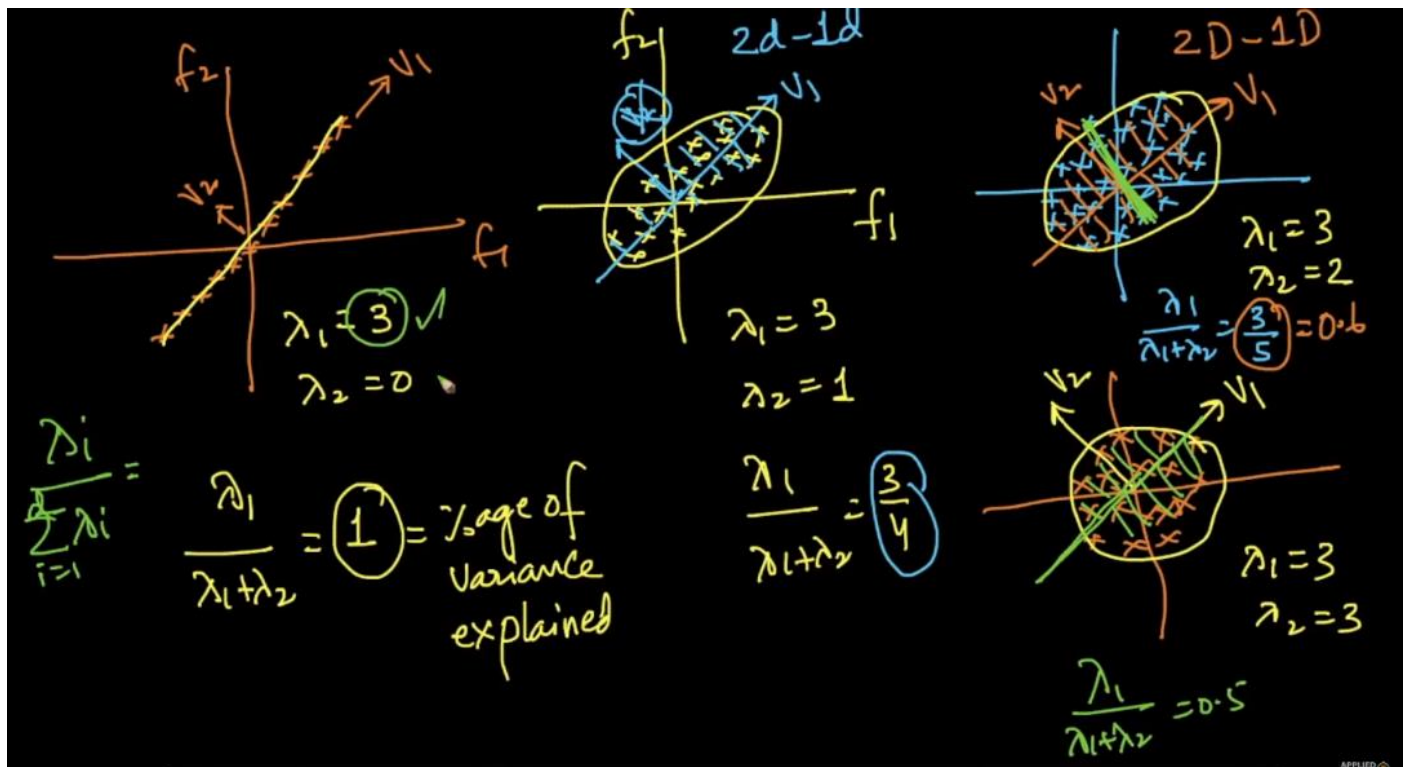


Given eigen values  $\lambda$ , we can say and find how much percentage of information will be retained by just taking new feature that is our new vector  $V_1$  or percentage of variance explained by

$$\lambda_1 / (\lambda_1 + \lambda_2).$$

in the first case where  $\lambda_1$  (wrt  $v_1$ ) = 3 and  $\lambda_2$  (wrt  $v_2$ ) = 0 i.e, means from 2d  $\Rightarrow$  1d if we simply ignore  $v_2$  still we retain 100% of information because  $\lambda_1 / (\lambda_1 + \lambda_2) \Rightarrow 3 / (3 + 0) \Rightarrow 100\%$

2<sup>nd</sup> case have  $\frac{3}{4}$  i.e 75%, that means  $V_1$  retains 75% of information, and so on.



## Comments:

1. Here the optimal unit vector  $u_1$  is used to indicate the direction of eigen vector  $v_1$  in which the maximum variance is preserved. We get the components of vector from  $v_1$  and it's direction is given by  $u_1$ . Only for indicating the direction of ' $v_1$ ', we use  $u_1$
2. Must read ref: <https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues>
3. Eigen faces and PCA in face detection: <https://sandipanweb.wordpress.com/2018/01/06/eigenfaces-and-a-simple-face-detector-with-pca-svd-in-python/>
4. Feature Scaling for PCA using sklearn: [https://scikit-learn.org/stable/auto\\_examples/preprocessing/plot\\_scaling\\_importance.html](https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html)
5. Applying PCA on iris dataset <https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60>