Introduction to Machine Learning Applications

Spring 2021

Lecture-10

Lydia Manikonda

manikl@rpi.edu

Rensselaer Polytechnic Institute · 1824 ·
Rensselaer

Today’s agenda

- Accuracy metrics
- Features and Dimensionality Reduction

Announcements

- Homework-4 due on March 4th 2021, 11:59 pm ET via LMS
- We will look at how to compute precision and recall values in a dynamic fashion.

Accuracy Metrics

		Actual Class	
		Class = 1	**Class = 0**
Predicted Class	**Class = 1**	f_{11}	f_{10}
	Class = 0	f_{01}	f_{00}

- f_{11} – True Positive
- f_{10} – False Positive – Type I error
- f_{01} – False Negative – Type II error
- f_{00} – True Negative

Precision: How many selected items are relevant?

Recall: How many relevant items are selected?

Precision

How many selected items are relevant?

$$Precision = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{retrieved\ documents\}|}$$

$$Precision = \frac{f_{11}}{(f_{10} + f_{11})}$$

Recall

How many relevant items are selected?

$$Recall = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{relevant\ documents\}|}$$

$$Recall = \frac{f_{11}}{(f_{01} + f_{11})}$$

F-measure

Better measure that considers the harmonic mean of *precision* and *recall*

$$f - measure = \frac{2*(precision*recall)}{(precision+recall)}$$

$$f1score = \frac{2 * precision * recall}{(precision + recall)}$$

Compute precision, recall and f-measure

tp = 8

fp = 4

fn = 2

tn = 6

		Actual Class	
		True	**False**
Predicted class	**True**	8	*4*
	False	2	6

Precision = tp/(tp+fp) = 8/(8+4) = 8/12

Recall = 8/(8+2) = 8/10

F-score = (2*precision*recall)/(precision+recall) = (2*8*8/120)/(8/10 + 8/12) =

Plot a precision recall graph

- True_labels = $[t_1, t_2, t_3, t_4, \ldots, t_k]$
- Pred_labels = $[p_1, p_2, p_3, p_4, \ldots, p_k]$

- At each point we compute the precision-recall values and dynamically compute them as we keep looking at all the labels consequently.

- Y_true = [1, 0, 0, 1, 1]
- Y_false = [1, 1, 1, 0, 0]

- i=2: precision, recall values

Actual = [1, 0, 0]

Pred = [1, 1, 1]

Tp=1

tn=0

Fp=2

Fn=0

Precision = 1/3 = 0.334

Recall = 1/1 = 1

- Y_true = [1, 0, 1, 0, 1]
- Y_false = [0, 1, 0, 1, 1]

- i=2: precision, recall values

Actual = [1, 0, 1]

Pred = [0, 1, 0]

Tp=0

tn=0

Fp=1

Fn=2

Precision = 0/1 = 0

Recall = 0/2 = 0

Sensitivity

Also considered as the True positive rate or equivalent to recall

$$Sensitivity = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{relevant\ documents\}|}$$

Specificity

Also known as True Negative Rate (actual negatives that are correctly identified)

$$Specificity = \frac{True\ Negative}{(True\ Negative + False\ Positive)}$$

$$Specificity = \frac{f_{00}}{(f_{00} + f_{10})}$$

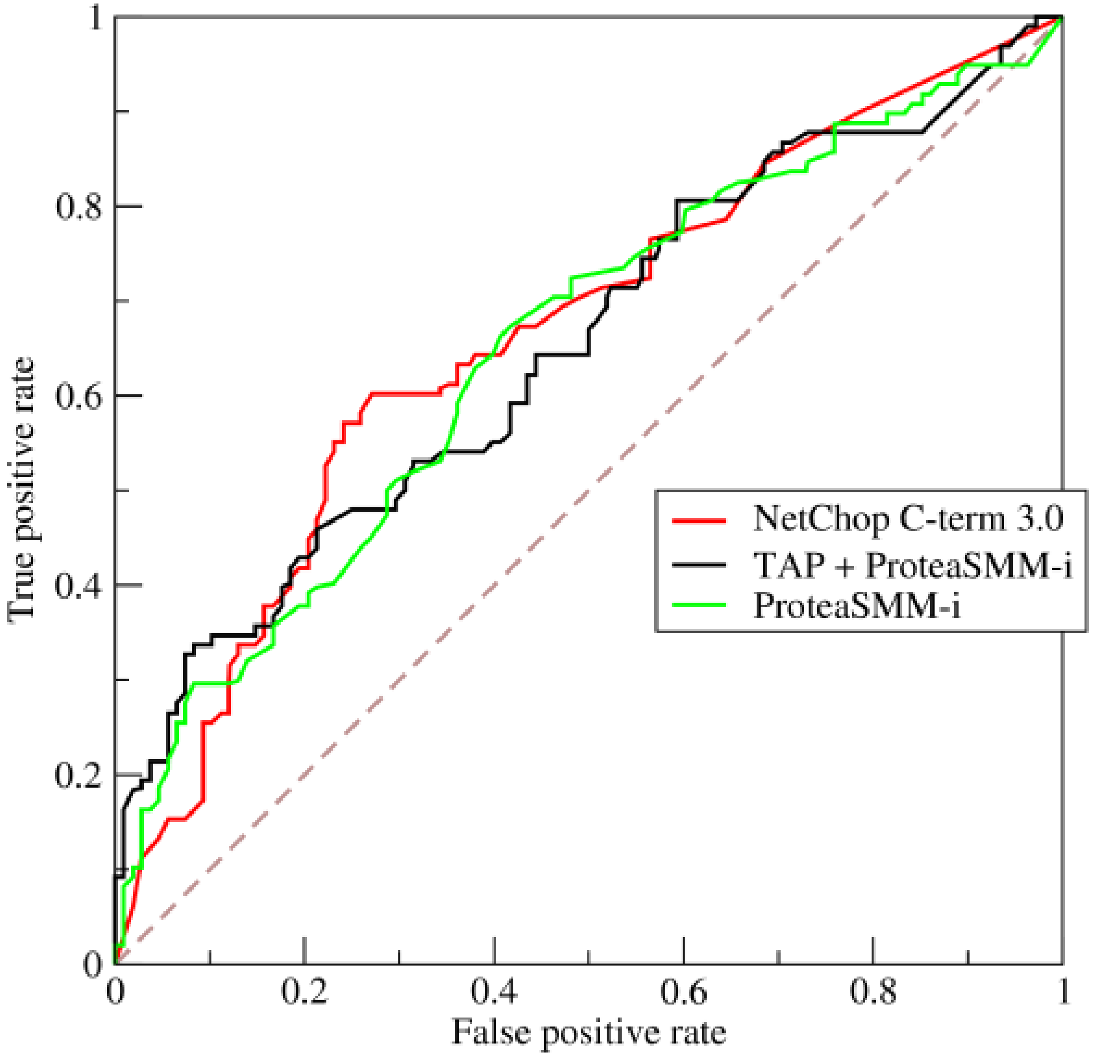
ROC curve

- Receiver operating characteristic curve
- Created by plotting True-positive rate vs False positive rate

$$FalsePositiveRate = \frac{False\ Positive}{(True\ Negative + False\ Positive)}$$

$$TruePositiveRate = \frac{True\ Positive}{(False\ Negative + True\ Positive)}$$

ROC curve – Example



https://upload.wikimedia.org/wikipedia/commons/6/6b/Roccurves.png

Class Exercises

		Actual Class	
		Cat (true)	**~Cat** (false)
Predicted Class	**Cat** (true)	5	*2*
	~Cat (false)	3	3

- Compute
  - Accuracy
  - Precision,
  - Recall,
  - f-measure,
  - Specificity
  - False positive rate

Accuracy = (5+3)/13

Tp=5

Fp = 2

Fn = 3

tn = 3

Precision = (5/5+2)

Recall = 5/(5+3)

F-measure = 2*p*r/(p+r)

$$\text{Accuracy} = \frac{TP + TN}{\#\text{Total}} = \frac{5 + 3}{5 + 2 + 3 + 3} = \frac{8}{13} = 61.53\%$$

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{5}{5 + 2} = \frac{5}{7} = 71.42\%$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{5}{5 + 3} = \frac{5}{8} = 62.5\%$$

$$\text{fscore} = \frac{2 \times Pre \times Rec}{(Pre + Rec)} = \frac{2 \times \frac{5}{7} \times \frac{5}{8}}{\left(\frac{5}{7} + \frac{5}{8}\right)} =$$

$$= \left(\frac{25}{28}\right) \Big/ \left(5 \times \left[\frac{15}{56}\right]\right)$$

$$= \frac{\cancel{25}}{\cancel{28}} \times \frac{\cancel{56}\,2}{\cancel{5} \times \cancel{15}\,3} = \frac{2}{3} = 66.7\%$$

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{3}{3 + 2} = \frac{3}{5} = 60\%$$

$$\text{FPR} = \frac{FP}{FP + TN} = \frac{2}{2 + 3} = \frac{2}{5} = 40\%$$

k-fold Cross-validation

- Resampling procedure to evaluate machine learning models on a given data sample.
- The parameter *k* refers to the number of groups that a given data sample is to be split into.
- If k=10, it is 10-fold cross-validation where the sample data is divided into 10 groups.

k-fold Cross-validation

> Shuffle the dataset (better)

> Split the dataset into *k* disjoint groups

> For each unique group:

   > Take the group as a hold out or test (validation) data set

   > Take the remaining groups as a training data set

   > Fit a model on the training set and evaluate it on the test set

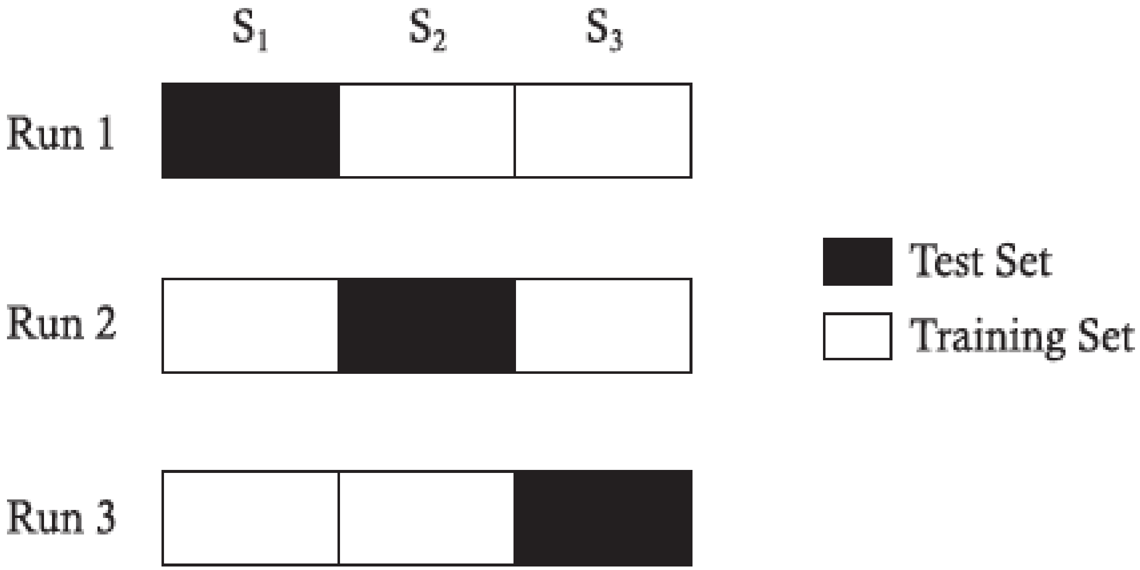
   > Record the evaluation score

> Find the mean of all the sample of model evaluation scores

k-fold Cross-validation

[1, 2, 3, 4, 5, 6]

Fold1: [5, 3]
Fold2: [1, 6]
Fold3: [2,4]

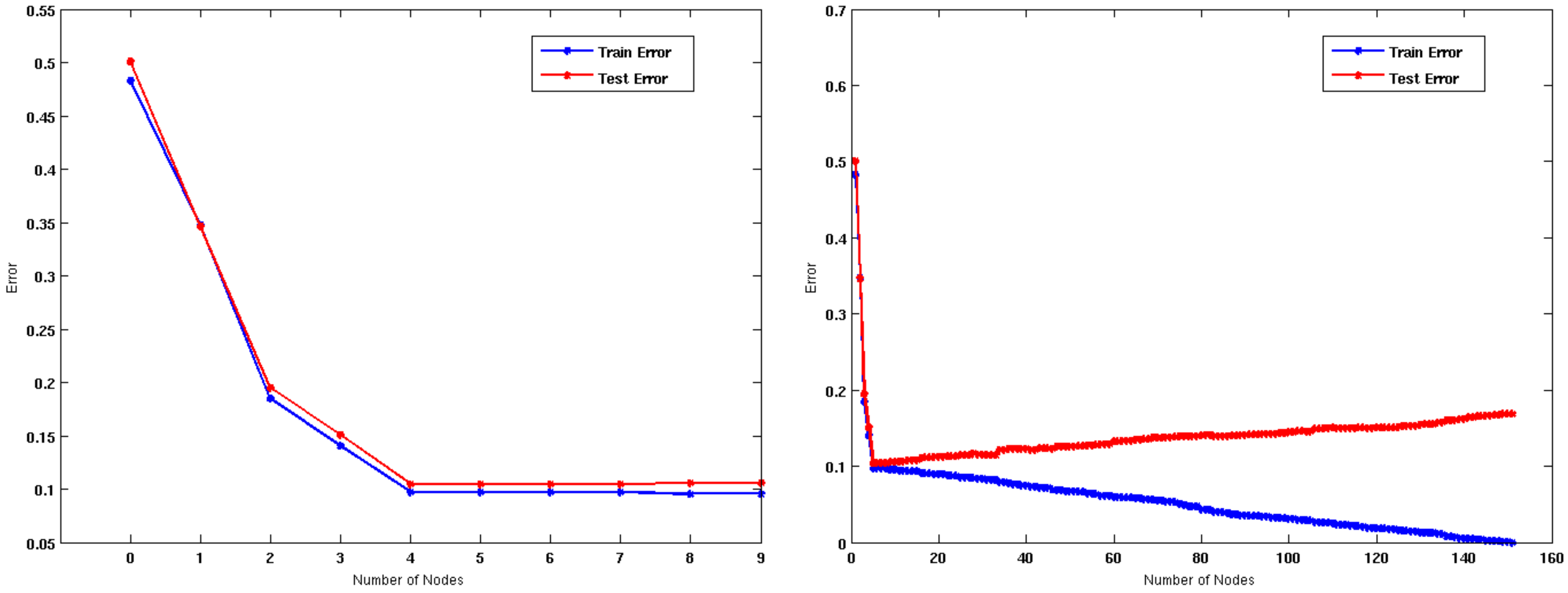


Model1: Trained on Fold2 + Fold3, Tested on Fold1
Model2: Trained on Fold1 + Fold3, Tested on Fold2
Model3: Trained on Fold1 + Fold2, Tested on Fold3

Example

- Given a set of data points – {a, b, c, d, e, f, g, h}
  - Perform 4-fold cross validation
  - Explain it in your own terms – what are the folds and how do you use them?

Model Overfitting & Underfitting



Underfitting: when model is too simple, both training and test errors are large

Overfitting: when model is too complex, training error is small but test error is large

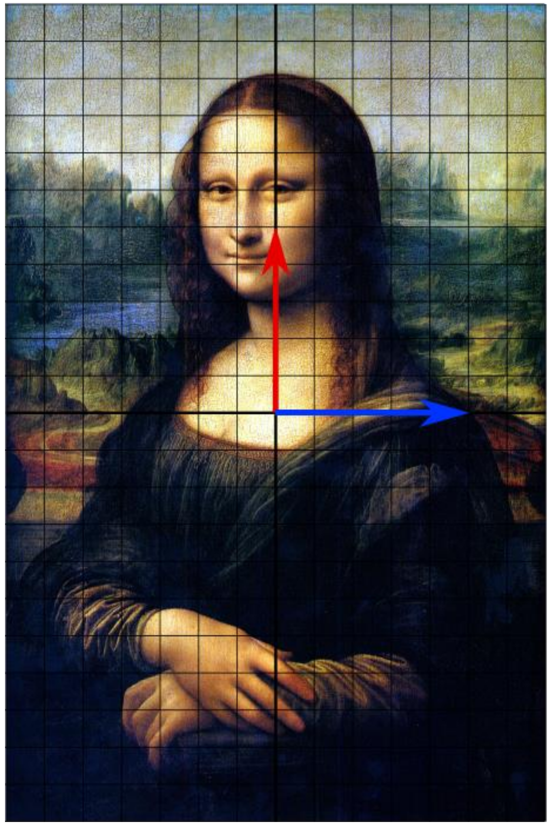
How do we convert ground truth data and predictions to a confusion matrix?

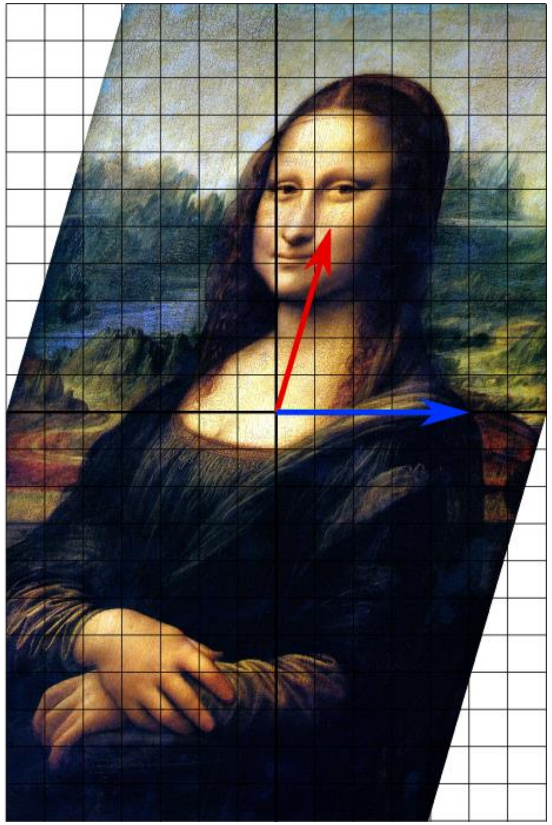
- Notebook example

- Exercises in the notebook to follow..

Eigenvalues and Eigenvectors

- In the image on the right, when the image is transformed, red arrow changed the direction. But the blue arrow didn’t – this is the eigenvector.
- Eigenvector does not change its direction.





Picture credits: By TreyGreer62 - Image:Mona Lisa-restored.jpg, CC0, https://commons.wikimedia.org/w/index.php?curid=12768508

Eigenvalues and Eigenvectors

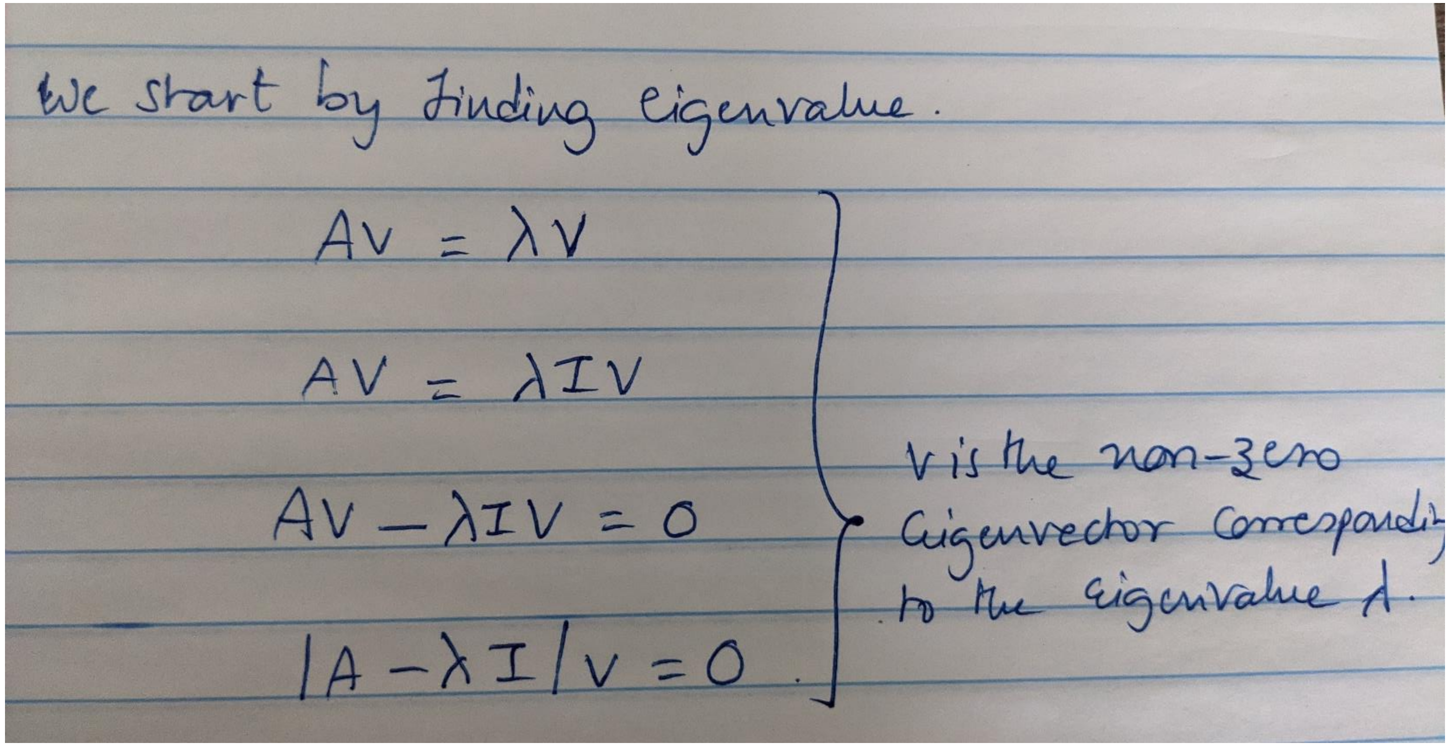
- Eigenvectors are the **characteristic vectors** that are nonzero vectors.

- Eigenvalues are the scalar values or **factors** with which corresponding eigenvectors are scaled.

- But how do we compute them?

Computing eigenvalues and eigenvectors

- We multiply a matrix with a vector and get the same result when we multiply a scalar by that vector.



Example: Computing eigenvalues and eigenvectors

If $A = \begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix}$, Compute eigenvalues and their Corresponding eigenvectors.

Start with: $|A - \lambda I| = 0$ ⟶ Finding the determinant.

$$\left| \begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = \left| \begin{bmatrix} -6-\lambda & 3-0 \\ 4-0 & 5-\lambda \end{bmatrix} \right| \quad \text{—} ①$$

$$\begin{vmatrix} -6-\lambda & 3 \\ 4 & 5-\lambda \end{vmatrix} = 0 \quad \text{—} ②$$

$$(-6-\lambda)(5-\lambda) - (3)(4) = 0 \quad \text{—} ③$$

$$-30 + 6\lambda - 5\lambda + \lambda^2 - 12 = 0 \quad \text{—} ④$$

$$\lambda^2 + \lambda - 42 = 0$$

$$(\lambda + 7)(\lambda - 6) = 0$$

$$\lambda = -7 \text{ or } 6.$$

We found eigenvalues.

Now compute corresponding eigenvectors

If $A = \begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix}$, Compute eigenvalues and their Corresponding eigenvectors.

Start with: $|A - \lambda I| = 0$ → Finding the determinant.

$$\left| \begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = \left| \begin{bmatrix} -6-\lambda & 3-0 \\ 4-0 & 5-\lambda \end{bmatrix} \right| \quad \text{— ①}$$

$$\begin{vmatrix} -6-\lambda & 3 \\ 4 & 5-\lambda \end{vmatrix} = 0 \quad \text{— ②}$$

$$(-6-\lambda)(5-\lambda) - (3)(4) = 0 \quad \text{— ③}$$

$$-30 + 6\lambda - 5\lambda + \lambda^2 - 12 = 0 \quad \text{— ④}$$

$$\lambda^2 + \lambda - 42 = 0$$

$$(\lambda + 7)(\lambda - 6) = 0$$

$$\lambda = -7 \text{ or } 6.$$

Case-1:
eigenvalue=6

Case 1: $\lambda = 6$:

$$Av = \lambda v$$

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 6 \begin{bmatrix} x \\ y \end{bmatrix} \quad (\to \text{Multiply})$$

$$\left.\begin{aligned} -6x + 3y &= 6x \\ 4x + 5y &= 6y \end{aligned}\right\} \text{——— (1)}$$

$$-12x + 3y = 0$$

$$4x - y = 0$$

$$\Downarrow$$

$$4x = y \quad \text{or} \quad y = 4x.$$

So, Eigenvector is any non-zero multiple of

$$\begin{bmatrix} 1 \\ 4 \end{bmatrix}.$$

Case-2: eigenvalue=-7

Case-2: $\lambda = -7$: $Av = \lambda v$

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (-7) \begin{bmatrix} x \\ y \end{bmatrix}$$

Multiplying these matrices:

$$-6x + 3y = -7x \quad \text{———} \; (1)$$

$$4x + 5y = -7y$$

$$x + 3y = 0 \quad \text{———} \; (2).$$

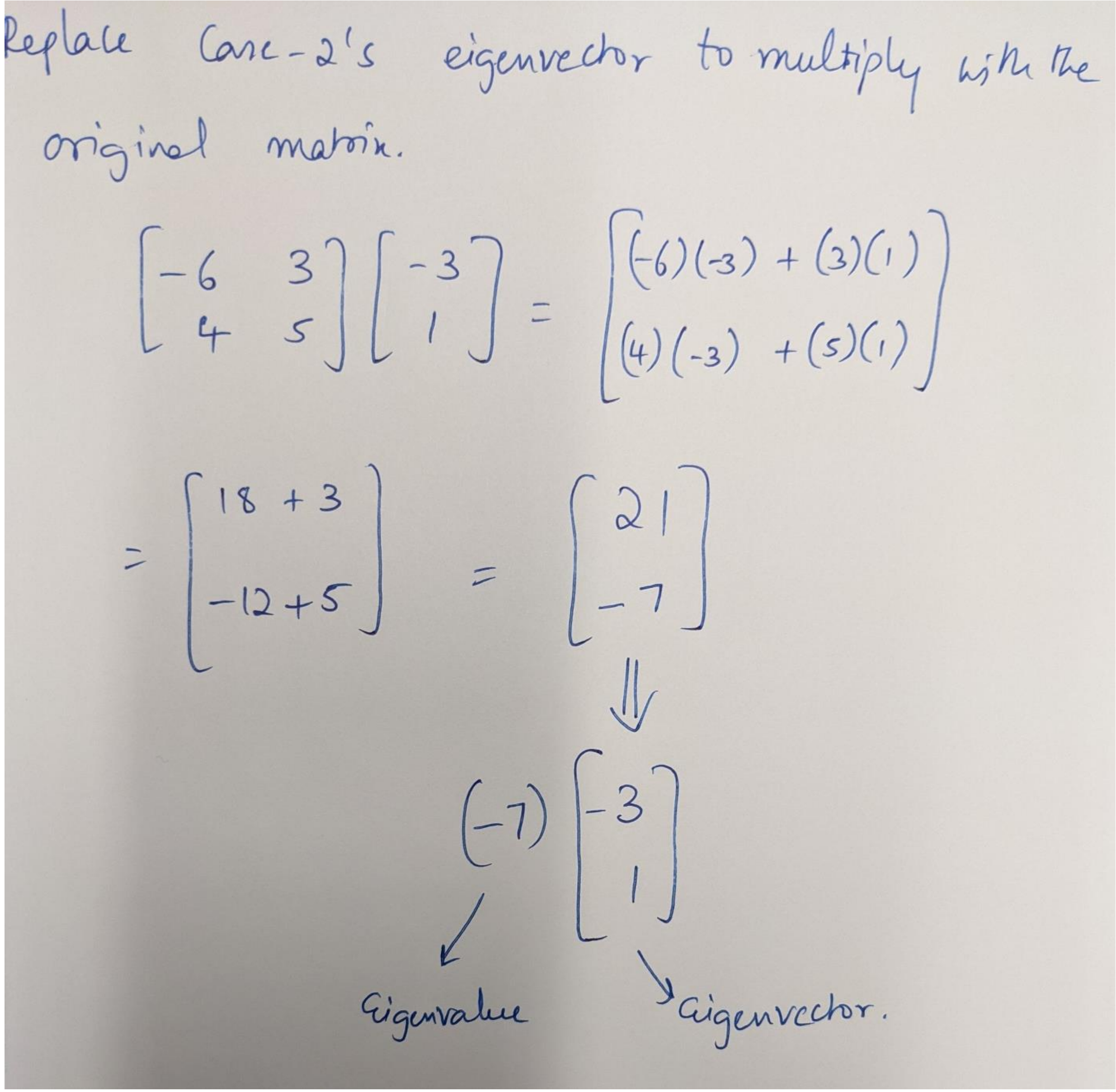
$$4x + 12y = 0$$

$$\Downarrow$$

$$x = -3y \quad \text{or} \quad y = \left(\frac{-1}{3}\right)x.$$

$$\begin{bmatrix} -3 \\ 1 \end{bmatrix}$$ → Eigenvector is any non-zero multiple of this vector.

Lets case-2's eigenvector and multiply with the original matrix



Example-2: eigenvalues and eigenvectors

Matrix is:

$$A = \begin{pmatrix} 2 & 2 \\ 5 & -1 \end{pmatrix}$$

Principal Component Analysis

- Step-1: Standardization
- Step-2: Compute covariance matrix
- Step-3: Compute the eigenvalues and eigenvectors of the covariance matrix
- Step-4: Sort the eigenvalues in a decreasing order
- Step-5: Choose the top-k eigenvectors which are the principal components – these will be the transformed feature vectors (reorient the data is the common approach)