Introduction to Machine Learning Applications

Spring 2021

Lecture-11

Lydia Manikonda

manikl@rpi.edu

Today’s agenda

- Features and Dimensionality Reduction
- Regression

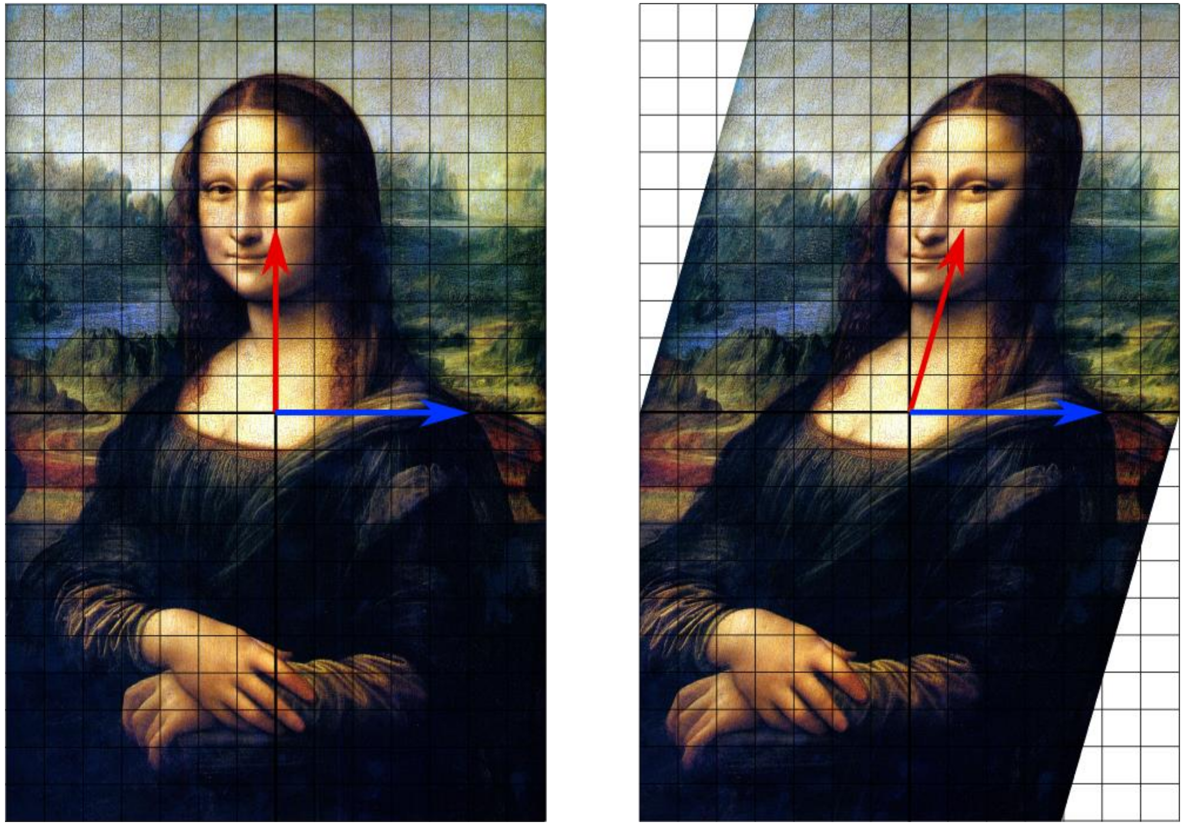
Announcements

- Homework-4 due on March 4th 2021, 11:59 pm ET via LMS

Dimensionality Reduction

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

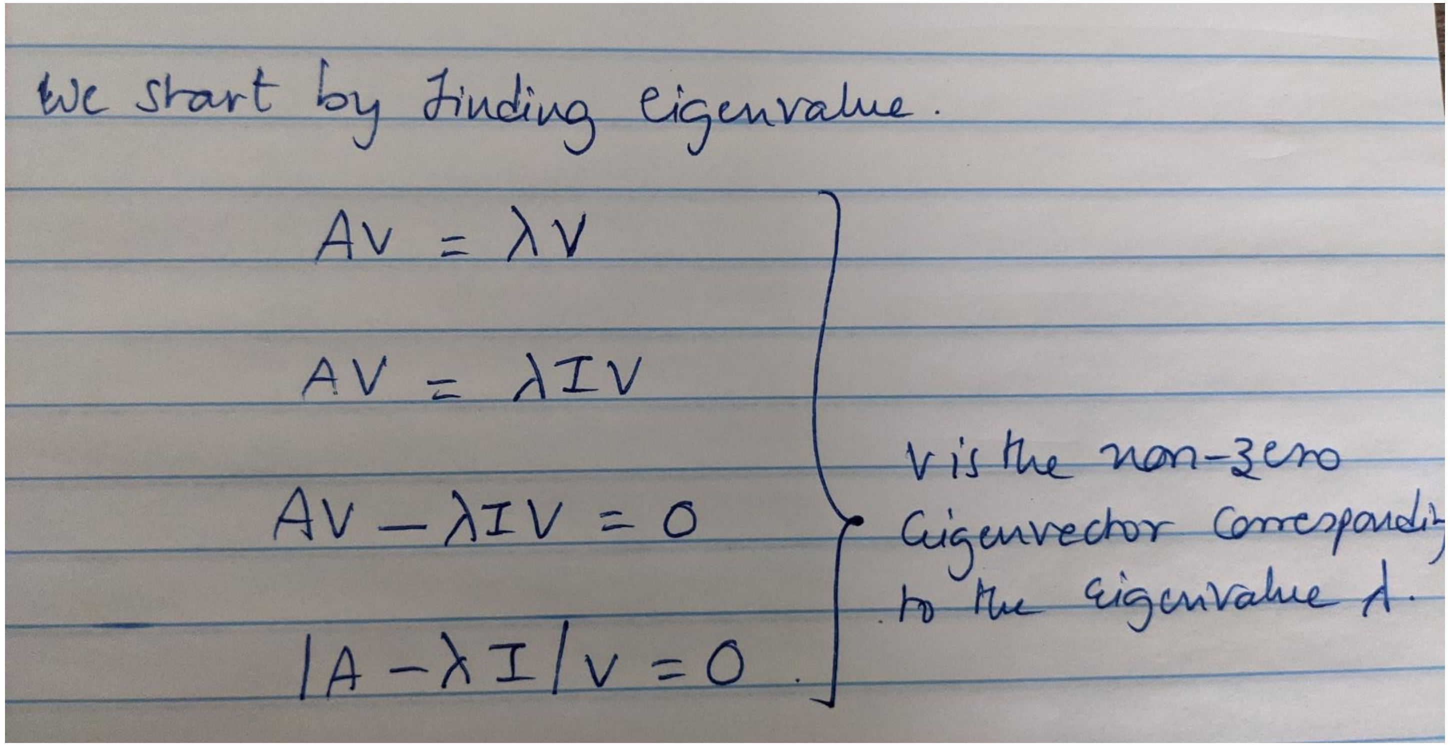
- Eigenvectors are the **characteristic vectors** that are nonzero vectors.

- 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.

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

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

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

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

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

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

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

We found eigenvalues.

Now compute corresponding eigenvectors

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

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: $\lambda = 6$:

$$Av = \lambda v$$

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

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

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

$$4x - y = 0$$

$$\Downarrow$$

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

So, Eigenvector is any non-zero multiple of

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

Case-2: eigenvalue=-7

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

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

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

Replace case-2's eigenvector to multiply with the original matrix.

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \begin{bmatrix} (-6)(-3) + (3)(1) \\ (4)(-3) + (5)(1) \end{bmatrix}$$

$$= \begin{bmatrix} 18 + 3 \\ -12 + 5 \end{bmatrix} = \begin{bmatrix} 21 \\ -7 \end{bmatrix}$$

$$\Downarrow$$

$$(-7) \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

(-7) → Eigenvalue

$\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ → Eigenvector.

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)

Regression

Linear Regression

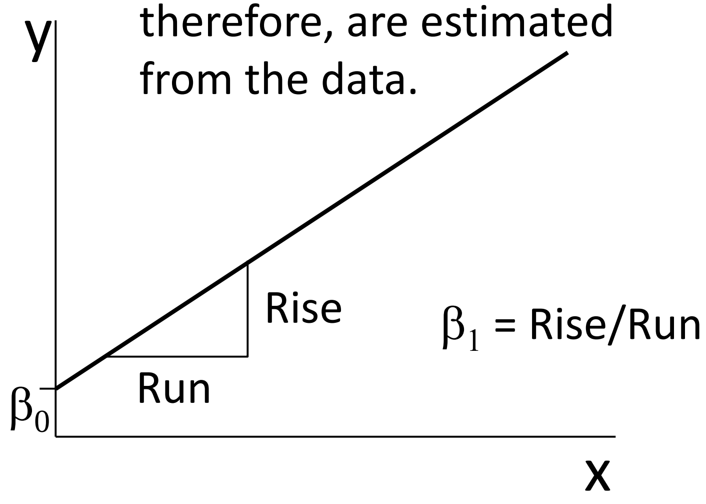
The technique is used to **predict** the value of one variable (the dependent variable - y) **based on** the value of other variables (independent variables x_1, x_2,…x_k) where ε is the error.

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Modeling

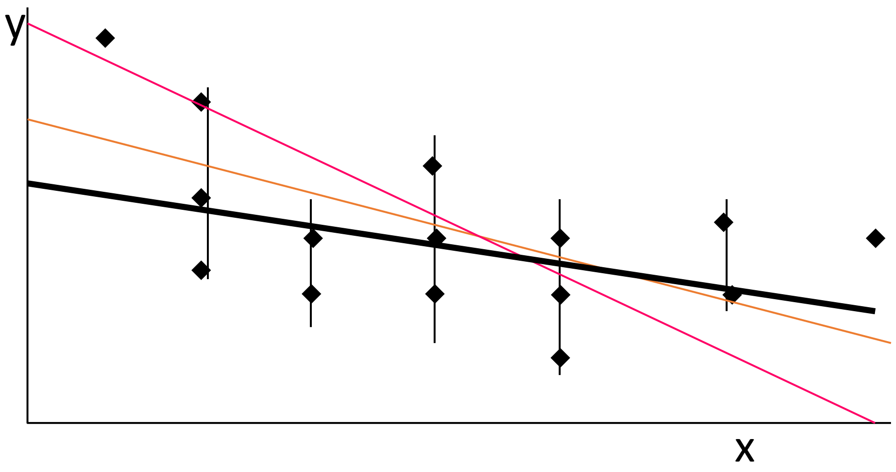
- The first order linear model

y = dependent variable
x = independent variable
β_0 = y-intercept
β_1 = slope of the line
$\mathcal{E}$ = error variable



Estimating the coefficients

- The estimates are determined by
  - drawing a sample from the population of interest,
  - calculating sample statistics.
  - producing a straight line that cuts into the data.



The question is:
Which straight line fits best?

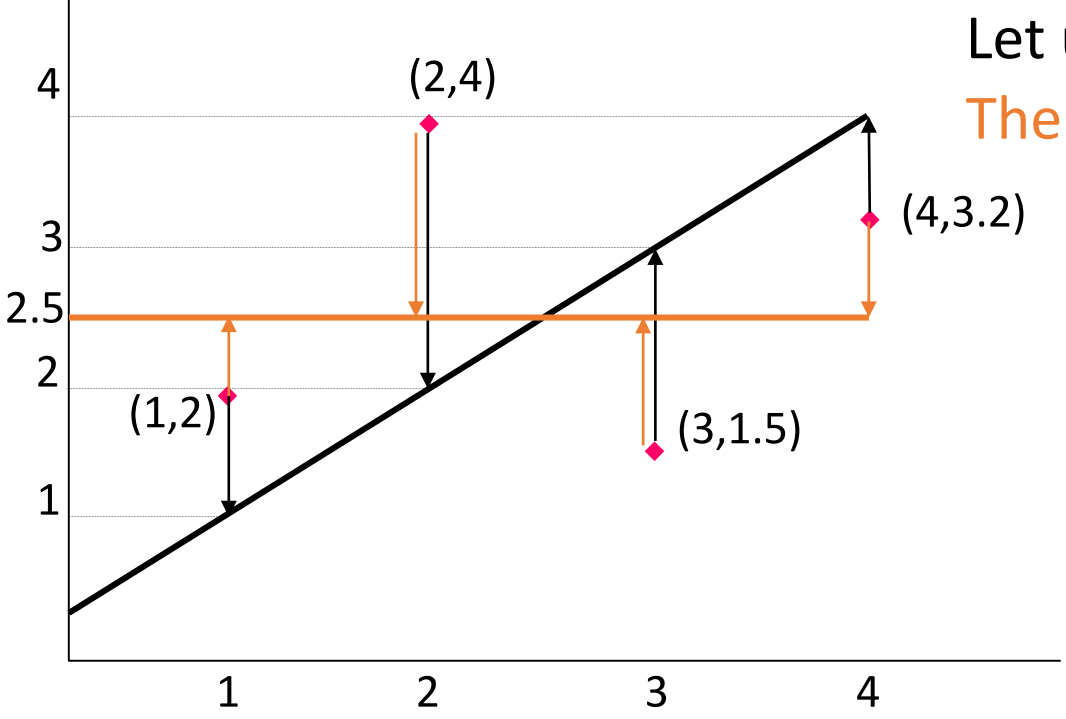
The best line is the one that minimizes the sum of squared vertical differences between the points and the line.

Sum of squared differences $= (2 - 1)^2 + (4 - 2)^2 + (1.5 - 3)^2 + (3.2 - 4)^2 = 6.89$

Sum of squared differences $= (2 - 2.5)^2 + (4 - 2.5)^2 + (1.5 - 2.5)^2 + (3.2 - 2.5)^2 = 3.99$

Let us compare two lines

The second line is horizontal



The smaller the sum of squared differences the better the fit of the line to the data.

Logistic Regression

- Special case of linear regression where the target variable is categorical in nature
- Uses a log of odds as a dependent variable
- Predicts the probability of occurrence of an event using a sigmoid function (inverse of logit function)

$$p = 1/(1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2)})$$

Linear vs Logistic Regression

- Output for linear regression is continuous
  - For example, stock prices
  - Or real estate price estimation
- Output for logistic regression is estimated as a constant
  - For example, predicting if a sample is tested +ve or –ve
  - Output >0.5 is +ve or 1 or yes; ouput <=0.5 is –ve or 0 or no

Linear vs Logistic Regression

- Linear regression is estimated using ordinary least squares
  - Distance minimizing approximation approach
  - Fits a regression line on a given set of data points that has the minimum sum of squared deviations (least squared error)

- Logistic regression is estimated using maximum likelihood estimation
  - “Likelihood” maximization method
  - Determines parameters (such as mean/variance) that are most likely to produce the set of data points.

Exercises -- Python notebook