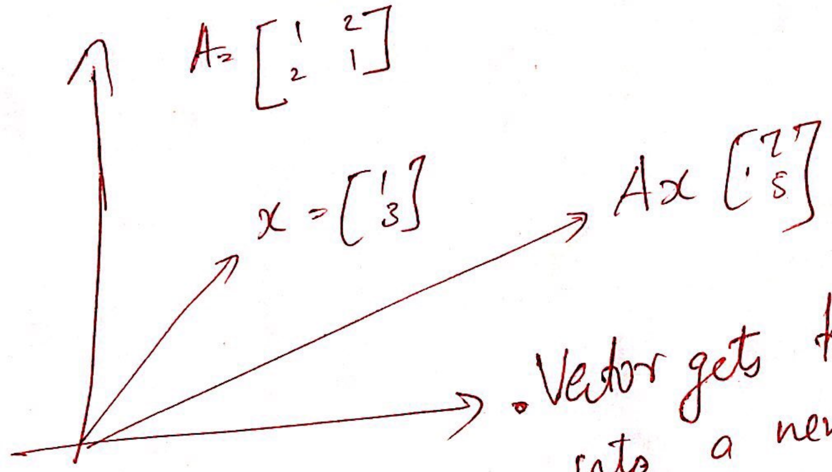
Lecture-6

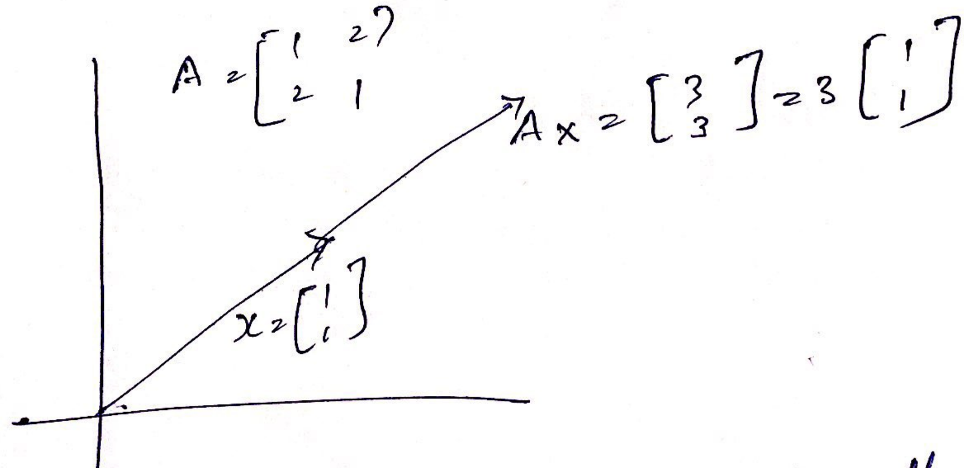
Module 6.1

Eigen Values and Eigen Vectors



- Vector gets transformed into a new vector.
- Vector may also be scaled in the process

Eigen



For a given square matrix A, there exist special vectors which refuse to stray from their path.

These vectors are called eigen vectors.

$Ax = \lambda x$

Chinese | Mexican

k_1 | k_2

$v_{(0)} = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$

p – fraction of students stay chinese in k_1.

$1-p$ – goes to mexican in k_1

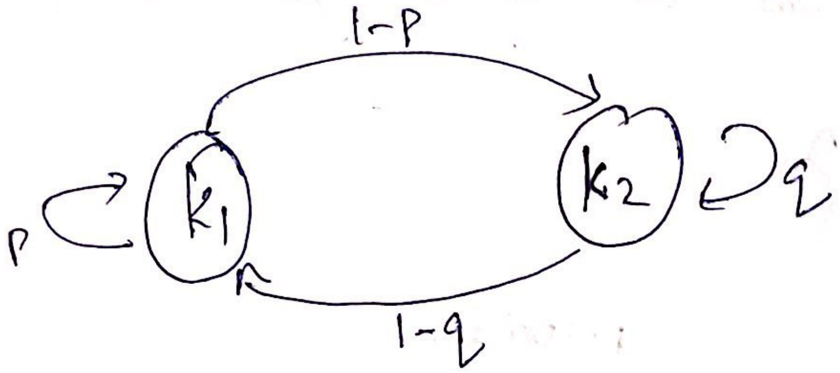
q – fraction of students stay mexican in k_2

$1-q$ – goes to chinese in k_2

$$v_{(1)} = \begin{bmatrix} pk_1 + (1-q)k_2 \\ qk_2 + (1-p)k_1 \end{bmatrix}$$

$$= \begin{bmatrix} p & 1-q \\ 1-p & q \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$$

$v_{(n)} = M^n v_{(0)}$



* Reach a steady state. ?

Let $\lambda_1, \lambda_2, \ldots \lambda_n$ be the eigenvectors of an $n \times n$ matrix A. λ_1 is called the dominant eigen value of A if

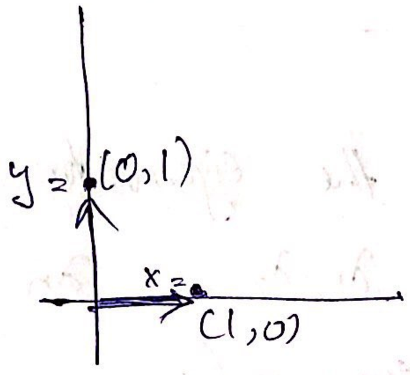
$$|\lambda_1| \geq |\lambda_i| \quad i = 2, \ldots, n$$

A matrix M is called a stochastic matrix if all the entries are positive and the sum of the elements in each column is equal to 1.

Theorem The largest (dominant) eigenvalue of a stochastic matrix is 1.

6.2

Linear Algebra - Basic Defenitions



Basis Vector

Fundamental vectors, using this we can express other vectors.

Gaussian Elimination → $O(n^3)$

Eigen Vectors can form a basis.

Eigen Vectors of a square symmetric matrix are even more special.

They form a very convenient basis.

Module
6.3

EigenValue Decomposition

$v_1, v_2, \ldots v_n$ be the eigenvectors of a matrix A. And $\lambda_1, \lambda_2 \ldots \lambda_n$ be the corresponding eigen values.

matrix U whose columns are $v_1, v_2 \ldots v_n$

$$AU = A \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & \cdots & v_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$= \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ Av_1 & Av_2 & \cdots & Av_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$Av_1 = \lambda_1 v_1$

$$= \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & \cdots & v_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & \cdots & & \lambda_n \end{bmatrix} = U\Lambda$$

$$AU = U\Lambda$$

If U^{-1} exists.

$$A = U\Lambda U^{-1}$$ [eigenvalue decomposition]

$$U^{-1}AU = \Lambda$$ [diagonalization of A]

U^{-1} exists if:

→ columns of U are linearly independent

→

If A is symmetric, situation is convenient.

$$Q = U^T U = \begin{bmatrix} \leftarrow u_1 \rightarrow \\ \leftarrow u_2 \rightarrow \\ \vdots \\ \leftarrow u_n \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow & & \uparrow & & \uparrow \\ u_1 & \cdots & u_2 & \cdots & u_n \\ \downarrow & & \downarrow & & \downarrow \end{bmatrix}$$

Q_{ij} is given by $u_i^T u_j$

$$Q_{ij} = u_i^T u_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

$\therefore U^T U = \mathbb{I}$ (the identity matrix)

Scanned with CamScanner

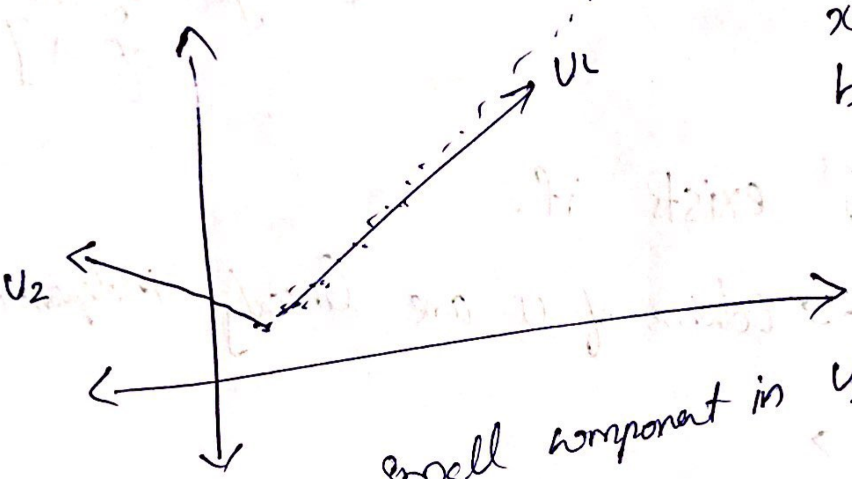
$$A = U \Sigma U^T$$

$$U \sum^{n} U^T x_0$$

Module 6.4

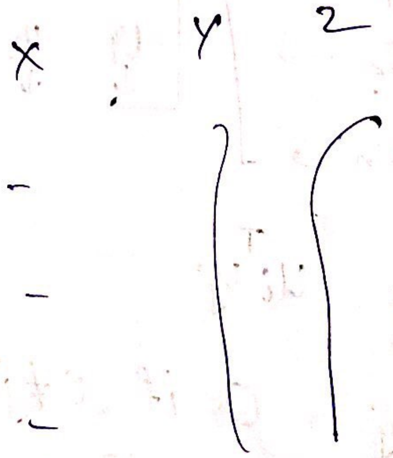
PCA & its interpretations.

We are using x and y as the basis



Small component in U_2 (noise)

Find the variance & throw away the lowest dimensions.



highly correlated.

Correlation

$$r_{yz} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(z_i - \bar{z})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^{n} (z_i - \bar{z})^2}}$$

Zero mean the data

P be $n \times n$ matrix such that $p_1, p_2 \dots p_n$ are the columns of P.

(UNIT VARIANCE & 0-MEAN)
↳ to the data.

$$x_i = \alpha_{i1} p_1 + \alpha_{i2} p_2 + \alpha_{i3} p_3 + \dots + \alpha_{in} p_n$$

For an orthogonal basis. α_i's using

$$\alpha_{ij} = x_i^T p_j = [\leftarrow x_i \rightarrow]^T \begin{bmatrix} \uparrow \\ p_j \\ \downarrow \end{bmatrix}$$

$$\hat{x_i} = [\leftarrow x_i^T \rightarrow] \begin{bmatrix} \uparrow & & \uparrow \\ p_1 & \dots & p_n \\ \downarrow & & \downarrow \end{bmatrix}$$

$$= x_i^T P_{n \times n}$$

$$= \alpha_{i1} \dots \alpha_{in}$$

$\hat{X} = XP$

$1^T X = 1^T X P = (1^T X) P$
$\hookrightarrow 0$

$1^T \hat{X} = 0.$

$X^T X$ is symmetric

Covarience Matrix
~~for~~

matrix $n \times n$

ij = Covarin b/w i^{th} j^{th}

x_{11} x_{21}

x_{12} x_{22}

$$\frac{1}{m} \sum_{i=1}^{m} (x_{1i} - \overbrace{\mu_1}^{\text{average}}) \cdot (x_{2i} - \mu_2) \quad x_{1m}$$

$\mu_1 = 0 = \mu_2$

$$\frac{1}{m} \sum_{i=1}^{m} x_{1i} \cdot x_{2i}$$ → dot product b/w the i^{th} column & the j^{th} column

$$C_{ij} = \frac{1}{m} \sum_{k=1}^{m} (x_{ki} - \mu_i)(x_{kj} - \mu_j)$$

$$= \frac{1}{m} \sum_{k=1}^{m} x_{ki} x_{kj}$$

$$= \frac{1}{m} X_i^T X_j = \frac{1}{m} (X^T X)_{ij}$$

$$\hat{X} = XP$$

$$\frac{1}{m} \hat{X}^T \hat{X} = \frac{1}{m} (XP)^T XP$$

$$= \frac{1}{m} P^T X^T X P$$

$$= P^T \left(\frac{1}{m} X^T X \right) P$$

$$= P^T \Sigma P$$

$$\frac{1}{m} (\hat{X}^T \hat{X}) = 0 \quad (i \neq j)$$

$$1/m (\hat{X}^T \hat{X}) \neq 0 \quad (i = j)$$

Scanned with CamScanner

$$\frac{1}{m}\hat{X}^T\hat{X} = P^T\Sigma P = D$$

Lecture 65 → PCA : Interpretation 2.

'n' orthogonal linearly independent vectors $P = p_1, p_2, \ldots p_n$, we can represent x_i exactly as a linear combination of these vectors

$$x_i = \sum_{j=1}^{n} \alpha_{ij} p_i$$ [we know how to estimate $\alpha_{ij}'s$] ~~but~~ ~~we~~ will ~~now~~

Interested in the top-k dimensions

$$\hat{x_i} = \sum_{j=1}^{k} \alpha_{ik} p_k$$

Select $p_i's$ such that we minimise the reconstructed error

$$e = \sum_{i=1}^{m} (x_i - \hat{x}_i)^2 \quad (x_i - \hat{x}_i)^T (x_i - \hat{x}_i)$$

$$e = \sum_{i=1}^{m} (x_i - \hat{x}_i)^T (x_i - \hat{x}_i)$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_{ij} p_j - \sum_{j=1}^{k} \alpha_{ij} p_j \right)^2$$

$$= \sum_{i=1}^{m} \left(\sum_{j=k+1}^{n} \alpha_{ij} p_j \right)^2 = \sum_{i=1}^{m} \left(\sum_{j=k+1}^{n} \alpha_{ij} p_j \right)^T \cdot \left(\sum_{j=k+1}^{n} \alpha_{ij} p_j \right)$$

$$= \sum_{i=1}^{m} \sum_{j=k+1}^{n} \alpha_{ij} p_j^T p_j \alpha_{ij} + \sum_{i=1}^{m} \sum_{j=k+1}^{n} \sum_{L=k+1, L \neq k}^{n} \alpha_{ij} p_j^T p_L \alpha_{iL}$$

$$= \sum_{i=1}^{m} \sum_{j=k+1}^{n} \alpha_{ij}^2$$

$$= \sum_{i=1}^{m} \sum_{j=k+1}^{n} (x_i^T p_j)^2$$

$$= \sum_{j=k+1}^{m} p_j^T \left(\sum_{i=1}^{m} x_i x_i^T \right) p_j$$

Lecture 6.6

Interpretation 3

6.7

Practical Example

- Consider a large image database

→ Each image is 100 X 100. [10 k dimension]

→ Store in much fewer dimension (50-200)

→ Construct a matrix $X \in \mathbb{R}^{m \times 10k}$

→ Each row corresponds to 1 image

→ Each image is represented using 10K dimensions

→ We retain the top 100 dimensions corresponding to the top 100 eigen vectors of $X^T X$

→ Eigen Faces

$$\sum_{i=1}^{1} \alpha_{1i} p_i$$

$$\sum_{i=2}^{2} \alpha_{1i} p_i$$

16 basis vectors

We will store the 'α'

6.8 Singular Value decomposition

Let $v_1, v_2 \ldots v_n$ be eigen vectors of A and let $\lambda_1, \lambda_2 \ldots \lambda_n$ be the corresponding values

$\cancel{Av_1}$ $Av_1 = \lambda_1 v_1$, $Av_2 = \lambda_2 v_2 \ldots$

$Av_n = \lambda_n v_n$

vector x in $\mathbb{R}^n$

$v_1, v_2 \ldots v_n$

$$x = \sum_{i=1}^{n} \alpha_i v_i$$

$$Ax = \sum_{i=1}^{n} \alpha_i A v_i = \sum_{i=1}^{n} \alpha_i \lambda_i v_i$$

(Rectangular Matrices are harder)

$A_{m \times n}$ transformation $\mathbb{R}^n \to \mathbb{R}^m$

Suppose, $\cancel{Ax_i = \sigma_i v_i}$

$$Av_i = \sigma u_i$$

$$x = \sum_{i=1}^{k} \alpha_i v_i$$

$$Ax = \sum_{i=1}^{k} \alpha_i A v_i = \sum_{i=1}^{k} \alpha_i \sigma u_i$$

$$Av_1 = \sigma_1 u_1 \;,\; Av_2 = \sigma_1 u_2 \;\cdots\; Av_k = \sigma_k u_k$$

$$A_{m \times n} V_{n \times k} = U_{m \times k} \Sigma_{k \times k}$$

$$A_{m\times n} V_{n\times n} = U_{m\times m} \Sigma_{m\times n}$$

$$U^T A V = \Sigma$$

$$[U^{-1} = U^T]$$

$$A = U \Sigma V^T \quad [V^{-1} = V^T]$$

$$A^T A = (U^T \Sigma V)^T (U^T \Sigma V^T)$$

$$= V \Sigma^T U^T U \Sigma V^T$$

$$A^T A = V^T \Sigma^2 V$$

$$AA^T = U^T \Sigma^2 U$$

$$A = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

$\sigma_i = \sqrt{\lambda_i}$ = singular value of A

U = left singular matrix of A

V = right singular matrix of A