⇒ PCA :-

↳ Eigenvectors → Vectors that just scale up/down on multiplication but never rotates.

↳ A $\boxed{\text{matrix}}$ can be represented in dimensions of eigenvectors

$N \times N$

Can have N eigenvectors. (orthogonal projections).

↳ Independent + somewhat interpretable.

↳ Influence of 1 dimension over others.

→ Translation independence

Eigen decomposition $A = V^T \Lambda V^T$

⇒ If eigenvalue = 0 → Singular matrix (As one of the vector useless).

⇒ $\beta^T X \beta X \beta +$

$L = (y - X\beta)^T (y - X\beta) \longrightarrow$ $[y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta]$

$\beta^T X^T X \beta - 2X^T y \beta$ → Quadratic Loss.

$\frac{\delta L}{\delta \beta} \longrightarrow -2X^T(y - X\beta)$ (Here at this β you will find a solution.)

→ If we find eigendecomposition of $X^T X$ i.e covariance, $\beta^T X (V^T \Lambda V) X\beta$, we will be able to speak something about the dist of β vectors (that are weights).

⇒ Additionally variability will be easily explained using eigenvalues of $X^T X$.

Singular Value Decomposition.

↳ same as Eigendecomposition

↳ + Mapping from $R^{M \times M}$ to $R^{N \times N}$ space (Two way)

→ Data space

↳ predictor space (Gi)

$A_{m \times n} = U D V^T$

U: $m \times m$ ↳ AA^T

V^T: ↳ $n \times n$ ↳ $A^T A$.

Symmetric Matrix

↳ Eigenvectors are orthogonal

⇒ Eigenvector of a matrix $A^{n \times n}$ → $A\hat{e} = T\hat{e}$

↳ There will be multiple projections of $\hat{e}$ for A

↳ Max → N.

↳ Proof:- Roots of x^n ↳ n. Max $[A - TI]$

↳ Can produce the same for. $X^T X$

↳ symmetric ∴ orthogonal eigenvalue

$Ax \cdot y = x \cdot Ay$. → dot product

⇒ New Story (NPTEL 6.1 IITM)

→ We have a vector V, A transformation A

Properties of A → column 1 or row sum = 1.

↳ symmetric

↳ Multiple times application of A ~~on~~ will generate a vector in direction of dominant eigenvector of A (Theorem).

Idea → $A \sim \dfrac{X^T X \text{ (corr)}}{\text{Variance (diagonal elems)}}$ → Correlation matrix
(Keep in Mind)

PCA → Linear Transformation + Unsupervised
Projections of given predictors/vectors to a new space,

↳ Reason or Objective: Maximize Variability

Data → For this optimization → $X^{n \times p}$

$\alpha_0 p_0 + \alpha_1 p_1 \ldots \alpha_{p-1} p_{p-1}$ → New projection
(Find α) that maximize ↳ Pro_1.

↳ $\dfrac{(Pro_1 - Mean(Pro_1))^2}{N}$ ↳ from the dataset

Dot product x_i with new axis → $\alpha_0 p_0 + \alpha_1 p_1 \ldots$

Optimization is Non-Linear
↳ Maybe Quadratic

Dec Vars → α's (projections)

↳ We will pass the dataset X as is. to the objective function
↳ Define what to minimize/optimize maximize

⇒ Not done yet → subtract this projection from the dataset and ~~find the~~ repeat the process again on the new data

↳ This is ~ to finding eigenvalues of $X^T X$ (covariance)

Largest

Eigenvalues of a transformation must be = 1 (for) stability

Ex - Corr (x, x)

$>1 \rightarrow$ Explode

$<1 \rightarrow$ Vanish.

N components.

Projections of PCA are basis

BASIS

$\because$ eigenvectors of corr mat.

vectors are not expressed as some of others.

In crux PCA$_i$'s are the basis in which we try to project our data

~~If basis~~, We need to find projections x_i's for this basis PCA$_i$

If orthogonal basis then projection can be easily found ~~pca~~

$\theta \rightarrow$ Eigenvectors are orthogonal

of a symmetric.

Linearly independent

Why we use this basis (multiple reasons:

1. PCA

$$\underset{n \times n}{AU} = A \overbrace{\begin{bmatrix} e_1 & e_2 & \dots & e_n \\ \downarrow & \downarrow & & \end{bmatrix}}^{U}_{n \times 1} = \begin{bmatrix} Ae_1 & Ae_2 & \dots & Ae_n \end{bmatrix}$$

$$\Leftarrow \begin{bmatrix} \lambda_1 e_1 & \lambda_2 e_2 & \dots & \lambda_n e_n \end{bmatrix}$$

$$\begin{bmatrix} e_1 & e_2 & e_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_n \end{bmatrix} = U \times \Lambda$$

$AU = U\Lambda$

$\therefore A = U \Lambda U^{-1}$

$Q \Rightarrow U^T U$ $= \begin{bmatrix} \leftarrow u_1 \rightarrow \\ \leftarrow u_n \rightarrow \end{bmatrix}_{n \times n} \begin{bmatrix} u_1 & & u_n \\ \downarrow & & | \end{bmatrix}_{n \times n}$

↳ eigenvector

Are orthogonal.

$\langle u_1 u_1 \rangle =$ ~~Some value~~ 1 ?

$U^T U_{ij} \Rightarrow \langle u_1 u_2 \rangle = 0$ ∴ $\langle u_i, u_j \rangle = 0$

$U^T U = I$ — 1 if $i=j$

Any of $A = U \Lambda U^T$; If A is iteratively on x_0 then final result → Stable if $\lambda = 1$
Exploding if $\lambda > 1$.

Let $A = N \times N$ matrix.
We wish to find max value of $x^T A x$ st $|x| = 1$.

$\mathcal{L} = x^T A x - \lambda(x^T x - 1)$ ∴ x must be an eigenvector of A with value λ

optimize $2Ax - \lambda 2x = 0$. $Ax = \lambda x$

* Why did we do this :-

So far → ◦ Different eigenvalues → diff eigenvecto
◦ Eigenvectors → Orthogonal ∴ Good basis

Unit Variance, Centered → helps in regularization (L1, L2)

Right eigenvalues for variance.

↳ Keep NN weights well regularized and applied.

Motivation of PCA :-

→ Reduce dimensions in which the data is well explained
→ project on new un-correlated dimensions

1. One data point x_i (~~m×n~~ R^n space) → # of predictors

$x_i = q_{i1}P_1 + q_{i2}P_2 \; -- \; q_{in}P_n$

→ P_1 is itself linear comb of P predictors

PCA is ~~fin~~ optimization to find P_i's. (1 to n)

~~Let~~ there be N data points
Then we need to find m new set of q_i's $_{1 \text{ to } N}$.

$\alpha_{1(1 \text{ to } N)} = \underset{1 \times N}{\langle x_1 \rangle} \times \underset{N \times N}{\begin{bmatrix} P_1 & P_2 & \cdots & P_n \end{bmatrix}} \rightarrow 1 \times N$

$\alpha_2 \qquad \langle x_2 \rangle \times \qquad \rightarrow 1 \times N$

$\vdots$

$\alpha_m \qquad \langle x_m \rangle \qquad 1 \times N$

Stack these together to find new projection map.

$\begin{matrix}\alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m\end{matrix} \rightarrow M \times N = XP$ → Let it be $\hat{X}$

→ eigenvectors. (Basis)

⇒ X^TX → is covariance (if X is zero centered
$m \times p$ or $m \times n$

~~Let say we have these~~. (Need to evaluate)

$\hat{X} = XP$

Covariance of the transformed data $= \hat{X}^T\hat{X}$

$\hat{X}^T\hat{X} = (XP)^T XP = P^T X^T X P = P^T (\Sigma) \times P$ ↳ covariance of the original data

↳ covariance of the transformed data

↳ Want this to be diagonal.

$AU = TU = UT$
$U^T A U = T$ — diagonal

We want $P^T \Sigma P = \text{Diag}$ ∴ P must be stack of eigenvectors of Σ

PCA → Linear transformation to the new basis where 1. variance is high and 2. cov is 0.

↳ Iteratively optimize

↳ Diagonalise the Σ

↳ This will give ~~stack of~~ eigenvectors as answer.

Any transformed data point

$x_i = \sum_{j=1}^{N} \alpha_{ij} p_j$ (We can remove some dimensions)

↳ Pure construction

m - data points

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

p - Reduced dimension

$\sum_{i=1}^{M} \left(\sum_{j=1}^{N} \alpha_{ij} p_j - \sum_{j=1}^{p} \alpha_{ij} p_j \right)^T \times (\quad)$

↳ $\sum_{i=1}^{m} \left(\sum_{j=p+1}^{N} (\alpha_{ij} p_j)^T (\quad) \right)$ $\quad p_i^T p_j = p_j^T p_i = 0 \quad i \neq j$

Same

$\sum_{i=1}^{m} [\alpha_{i,p+1} p_{p+1} + \alpha_{i,p+2} p_{p+2} \ldots \alpha_{iN} p_N)^T (\quad)$

$$= \sum_{i}^{M} \sum_{j=k+1}^{N} (q_{ij})^2 \quad = \sum_{i} \sum_{j=k+1}^{N} (x_i^T p_j)^2$$

$$= p_j^T \sum_{i=1}^{M} x_i^T \vec{x}_i \; p_j^{T}$$

↳ covariance

Minimize $p_j^T \Sigma p_j$

↳ Find smallest eigenvalues of Σ

↳ Throw away these dimensions

PCA - orthogonal basis (so that the ~~q_i~~ projection are easily found)

Covariance → Eigenvectors of covariance

Property - ← ~~⟸~~ Stack of these e_i's

↳ cov of transform = Diagonal.

gives a

Eigenvalues ← ~~eigenvectors~~ orthogonal basis.

can also be used for dimensionality reduction.

↳ Transformed projection can be easily computed

Take dot product of datpoint with basis/eigenvectors.

Bcos eigenvectors are orthogonal. $P = [e_1 \, e_2 \, e_n$

↳ as $X^T X$ is symmetric

$X_{transformed} = X \cdot P = \langle X, P \rangle$

↳ [eigenvectors]

$X_{transformed_i} = X \cdot P_i = \langle X, P_i \rangle$

$$\text{Variance} = X_{tra}^T \times X_{transformed}$$

$$= (X P_i)^T (X \cdot P_i) = P_i^T (X^T X P_i) \quad \rightarrow \text{eigenvector}$$

$$\Rightarrow P_i^T \times \lambda_i P_i = \lambda_i P_i^T P_i = \lambda_i$$

To maximize variance select the largest P_i

Example :-

↳ Image (100 × 100) → 10K dimensions

↓ Need to compress the data

50 - 200 dimensions.

$X \rightarrow M \times 10K$.

↳ $X^T X$ → Then compute the eigendecomposition of the same

↳ Each eigenvector is represented in 10K dimensions.

↳ Linear combination of original features.

Scanned with CamScanner

We have 10000 eigenvectors

Each eigenvector → 1 map of 100 × 100 pixel

$x_i x p_1$ → 1st Map.

Original Image (i^{th})

The difference between $(x_i - x_i p_1)^2$ → Reconstruction error.

Only 20-50 basis p_i's (top ones) will help us explain the image

⇒ Originally 5 ~~1000~~ images. ~ 100 × 100 pixel each.

⇒ We stored. 10K features for 5 images [5 × (10K)]

Each represented in 10K dims.

Now we only need 20 eigenvectors (basis)

↳ If we have q's for 20 basis.

↳ [20 × (10K)] → This is transformation issue.

But we can now explain 5 images in just 20 dimensions

(Transformat थोड़ झेलना पड़ेगा, dimensions will reduce)

PCA → Deals with data Augmentation

↳ Remove less significant dimensions.

explainability (Noise)

Data Augmentation

↳ Lighting, Rotation, Translation

↳ Creates noise base variation

↳ Can be removed.