(*) SVD :-

→ From the singular value decomposition of matrix A. we can get matrix B ~~which~~ of rank K which Best approximates A.

→ columns of V ← right singular vector of A. } orthognal.
columns of U ← left singular vector of A }

$A_{n \times d}$ ← dimension.
↑
no. of data points

∘ objective :. Find best K-dimensional subspace w.r.t set of points.

Here, best means minimize the sum of the squares of the $\perp^{er}$ distances of the points to the subspace.

$$\boxed{A_{m\times n} = U_{m\times r}\,\Sigma_{r\times r}\,V^T_{r\times n}}$$

$$\boxed{A \approx U\Sigma V^T = \textstyle\sum_i \sigma_i\, u_i \cdot v_i^T}$$

scalar vectors.

A: Input Data matrix
(m×n : m ← no. of doc. , n ← no. of terms)

U: Left singular vectors.
(m×r : m ← doc , r ← concepts)

Σ: singular values. (Diagonal matrix.)
(r×r : strength of each concept)
(r: rank of matrix. A)

V: Right singular matrix
(n×r : n ← terms , r ← concepts)

U,V: column are orthonormal,
$U^TU = V^TV = I$.

Σ: Diagonal
– Entries are +ve.
$(\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq 0)$

Applications of SVD :-

1) Low rank approximation.
2) Signal and image Analysis.
3) pseudo inverse and Least square approximation.

$$A_{m\times n} = U_{m\times n} S_{n\times n} V^T_{n\times n}$$ } Reduced SVD.

$$A_{m\times n} = U_{m\times m} S_{m\times n} V^T_{n\times n}$$ } Full SVD

U and V are orthogonal ~~and S is dia~~

S is diagonal matrix.

Here, $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_n$ (non-Increasing order).

σ_i's are singular values (non-negative + real)

*) How to find matrix U?

$A.A^T_{(m\times m)}$

→ columns of U are orthonormal eigen vectors of AA^T.

→ columns of V are orthonormal eigen vectors of A^TA.

$$A.A^T = USV^T(VS^TU^T)$$
$$= USV^TVS^TU^T$$
$$= USS^TU^T \quad [V^TV = I]$$
$$= U\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ \vdots & \sigma_2^2 & 0 \\ \vdots & 0 & \sigma_n^2 \end{bmatrix} U^T$$

→ singular values are the square root of eigen values of $(A.A^T)$ or (A^TA).

(*) MLE of Poisson Distribution :-

Find the maximum likelihood estimate for the parameter λ. for a random sample $x_1, x_2, \ldots x_n$. of a poisson distribution.

Also find its variance.

Our Target

Find the MLE of λ

$\rightarrow$ For poisson distribution we have

PMF of Poisson Distrib^n

$$f(x_i) = \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!} \quad ; \quad x_i = 0, 1, 2, \ldots$$

The Likelihood function L defined as.

$$L = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!}$$

Taking log likelihood,

$$\log L = \sum_{i=1}^{n} \log\left(\frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!}\right)$$

$$= \sum_{i=1}^{n} \log(e^{-\lambda}) + \log(\lambda^{x_i}) - \log(x_i!)$$

$$= \sum_{i=1}^{n} -\lambda \log e + x_i \log\lambda - \log(x_i!)$$

$$= -\lambda n + \log\lambda \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log(x_i!)$$

$$\frac{d}{d\lambda} \log L = 0$$

$$\Rightarrow -n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i = 0$$

$$\therefore \boxed{\lambda = \frac{\sum_{i=1}^{n} x_i}{n}} \qquad \lambda = \bar{x}$$

$$\frac{d^2}{d\lambda^2} \log L = -\frac{1}{\lambda^2} \sum_{i=1}^{n} x_i < 0$$

Hence, the MLE for λ is sample mean $\bar{x}$.

The variance of the estimate is given by

$$\frac{1}{V(\lambda)} = E\left\{\frac{-\partial^2}{\partial\lambda^2}\log L\right\}$$

$$= E\left\{\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i\right\}$$

$$= E\left\{\frac{n\bar{x}}{\lambda^2}\right\}$$

$$= \frac{n}{\lambda^2} E\{\bar{x}\}$$

$$= \frac{n}{\lambda^2} \cdot \lambda = \frac{n}{\lambda}$$

$$\frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

$$\therefore \sum_{i=1}^{n} x_i = n\bar{x}$$

$$\boxed{Var(\lambda) = \frac{\lambda}{n}}$$

(*) MLE of Geometric Distribution :-

Our Target :- find MLE of p.

unknown parameter :- p.

For Geometric Distribution, we have

Here, $q = (1-p)$

$$f(x_i) = q^{x_i - 1} p \quad ; \; x_i = 1, 2, \ldots$$

$$\therefore f(x_i) = (1-p)^{x_i - 1} p.$$

The Likelihood function L is defined as,

$$L = \prod_{i=1}^{n} f(x_i | \theta)$$

$$= \prod_{i=1}^{n} (1-p)^{x_i - 1} p$$

Taking log likelihood,

$$\log L = \sum_{i=1}^{n} \log (1-p)^{x_i - 1} p.$$

$$\log L = \sum_{i=1}^{n} \log (1-p)^{x_i - 1} + \log p.$$

$$\log L = \log(1-p)\sum_{i=1}^{n}(x_i - 1) + \sum_{i=1}^{n}\log p$$

$$\log L = \log(1-p)\left(\sum_{i=1}^{n} x_i - n\right) + n\log p$$

$$\frac{d}{dp}\log L = 0$$

$$\Rightarrow \frac{-1}{(1-p)}\left(\sum_{i}^{n} x_i - n\right) + \frac{n}{p} = 0$$

$$\frac{n}{p} = \frac{\sum x_i - n}{(1-p)}$$

$$n - \cancel{np} = p\sum x_i - \cancel{np}$$

$$p = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

$$\boxed{p = \frac{1}{\bar{x}}}$$

$$p_{MLE} = \frac{1}{\bar{x}}$$

(*) MLE for the Bernouli Distribution :-

we have Bernoulli Distribution,

$$f(x_i) = p^{x_i}(1-p)^{1-x_i} \; ; \; x_i = 1, 2, \ldots$$

The Likelihood function defined as.

$$L = \prod_{i=1}^{n} f(x_i|\theta)$$

[∵ assuming all observation are independent to each other]

$$L = \prod_{i=1}^{n} p^{x_i}(1-p)^{(1-x_i)}$$

Taking Log of the likelihood,

$$\log L = \sum_{i=1}^{n} \log\left(p^{x_i} \cdot (1-p)^{(1-x_i)}\right)$$

$$\log L = \sum_{i=1}^{n} x_i \log p + \sum_{i=1}^{n} (1-x_i)\log(1-p)$$

$$\log L = \log p \sum_{i=1}^{n} x_i + \log(1-p)\left(n - \sum_{i=1}^{n} x_i\right)$$

Estimating Likelihood equation,

$$\frac{d}{dp} \log L = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{(1-p)}\left(n - \sum_{i=1}^{n} x_i\right) = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{p} = \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$\sum_{i=1}^{n} x_i - \cancel{\left(\sum_{i=1}^{n} x_i\right) p} = np - \cancel{p\sum_{i=1}^{n} x_i}$$

$$p = \frac{\sum_{i=1}^{n} x_i}{n}$$

(*) MLE of Exponential Distribution :-

Our Target :-
To find MLE of λ.

Exponential Distribution,

$$f(x) = \lambda e^{-\lambda x} \; ; \; x \geq 0$$

The Likelihood function L is defined as,

$$L = \prod_{i=1}^{n} f(x_i|\theta)$$

$$L = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

Taking log likelihood,

$$\log L = \sum_{i=1}^{n} \log(\lambda e^{-\lambda x_i})$$

$$= \sum_{i=1}^{n} \log \lambda + \log e^{-\lambda x_i}$$

$$= n \log \lambda - \sum_{i=1}^{n} \lambda x_i$$

$$\frac{d}{d\lambda}(\log L) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\frac{d^2}{d\lambda^2} \log L = \frac{-n}{\lambda^2} < 0$$

$$\therefore \frac{n}{\lambda} = \sum_{i=1}^{n} x_i$$

$$\therefore \lambda = \frac{n}{\sum_{i=1}^{n} x_i}$$

$$\therefore \lambda = \frac{1}{\bar{x}}$$

$\therefore \lambda = \frac{1}{\bar{x}}$ is max

$$\therefore \boxed{\text{MLE of } \lambda = \frac{1}{\bar{x}}}$$

(*) MLE for Binomial Distribution :-

For Binomial Distribution we have,

$$f(x_i) = \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i} \quad ; x_i = 1,2,3,\ldots n \quad p \in [0,1]$$

The Likelihood function is defined as,

$$L = \prod_{i=1}^{n} f(x_i|\theta)$$

$$L = \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i} (1-p)^{(n-x_i)}$$

Taking log of the likelihood L,

$$\text{Log } L = \sum_{i=1}^{n} \log\left(\binom{n}{x_i} \cdot p^{(x_i)} \cdot (1-p)^{(n-x_i)}\right)$$

$$\text{Log } L = \sum_{i=1}^{n} \left(\log\binom{n}{x_i} + x_i \log p + (n-x_i) \log(1-p)\right)$$

$$\text{Log } L = \sum_{i=1}^{n} \log\binom{n}{x_i} + \log p \sum_{i=1}^{n} x_i + \log(1-p) \sum_{i=1}^{n} (n-x_i)$$

$$\frac{d}{dp} \log L = 0$$

$$\Rightarrow 0 + \frac{1}{p} \Sigma x_i - \frac{1}{(1-p)} \sum_{i=1}^{n} (n-x_i) = 0$$

$$\frac{\Sigma x_i}{p} = \frac{n^2 - \sum_{i=1}^{n} x_i}{(1-p)}$$

$$\frac{\Sigma x_i}{n^2} = p$$

(*) Fisher's LDA :-

⇒ Learning set is labeled ← supervised learning.

→ Unlike the PCA having unlabelled data

→ In PCA we were interested in the direction in which maximum scatter of the entire data exist. whereas in LDA we want to maximize class separability.

→ select W to maximize the ratio of between class scatter and within-class scatter.

Between-class scatter matrix is defined by-

$$S_B = \sum_{i=1}^{c} N_i(\mu_i-\mu)(\mu_i-\mu)^T$$

μ_i ← mean of class X_i
N_i ← no. of samples in X_i

within-class scatter matrix is defined by-

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k-\mu_i)(x_k-\mu_i)^T$$

$$\boxed{W_{opt} = \text{argmax} \frac{|W^T S_B W|}{|W^T S_W W|}}$$

$W_{opt} = [w_1, w_2, \dots, w_m]$

$\{w_i | i=1,2,3,\dots,m\}$ is the set of eigenvectors of $\boxed{S_W^{-1} S_B}$

$$\boxed{S_B w_i = \lambda_i S_W w_i}$$

→ There are atmost $(C-1)$ non-zero eigen values. So, upper bound of m is $(C-1)$.

→ S_W is singular if $N<D$. It's rank is ~~atmost~~ atmost $(N-C)$.

Solution:
- Project the samples to a lower dimensional space of the feature space
- use PCA to reduce dimension to $(N-C)$.
- Then apply standard FLD to reduce dimension to $(C-1)$

$\Rightarrow$ W_{opt} given by $\boxed{W_{opt} = W_{fld}^T W_{pca}^T}$

$$\boxed{W_{pca} = \underset{W}{\operatorname{argmax}} \left| W^T S_{\cancel{T}} W \right|}$$

$(S_W + S_B)$

$$\boxed{W_{fld} = \underset{W}{\operatorname{argmax}} \frac{\left| W^T W_{pca}^T S_B W_{pca} W \right|}{\left| W^T W_{pca}^T S_W W_{pca} W \right|}}$$