Inference & Representation
Recitation #1

Today's Agenda :

1. Review of concepts needed to understand PCA
2. Gaussian Random Variable / Vector
3. KL-Divergence

1.1. Diagonalization

<Def.> $A \in \mathbb{R}^{n \times n}$, $A = Q \Lambda Q^{-1}$

* A diagonalizable if $\exists\ v_1 \cdots v_n$ independent eigenvectors

  <Recall> $v \in \mathbb{R}^n$ eigenvector if $\exists\ \lambda \in \mathbb{R}$ s.t. $Av = \lambda v$

* Q is said to diagonal A.

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \quad Q = [\, v_1 \mid \cdots \mid v_n \,]$$

(Ex.) $A = \begin{bmatrix} -3 & -4 \\ 5 & 6 \end{bmatrix}$

Eigenvalues of A : $\det(A - \lambda I) = 0 \Rightarrow \lambda_1 = 1,\ \lambda_2 = 2$

Eigenvectors of A : $v_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $v_2 = \begin{bmatrix} -0.8 \\ 1 \end{bmatrix}$

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad Q = \begin{bmatrix} -1 & -0.8 \\ 1 & 1 \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} -5 & -4 \\ 5 & 5 \end{bmatrix}$$

* Some metrices are NOT diagonalizable :
when # L.I. eigenvectors < dim(A)

(Ex.) $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ eigenvalues : $\det(A - \lambda I) = 0 \Rightarrow \lambda_1 = \lambda_2 = 1$

multiplicity $= 2$

eigenvectors : $v_1 = v_2 = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$ $\updownarrow\times$

$\dim(\text{Ker}(A)) = 1$

So, A : Non-diagonalizable matrix

1.2. Spectral Theorem

<Recall> Symmetric : $A = A^T$
Orthogonal : $VV^T = V^TV = I$

<Def.> $A \in \mathbb{R}^{n\times n}$ diagonalizable $(A = Q\Lambda Q^{-1})$, Q can be chosen orthogonal
$\exists$ orthonormal basis of $\mathbb{R}^n$ compose eigenvectors of A

<Def.> $A = PDP^T$ P: orthogonal $= [v_1 | \cdots | v_n]$

D: diagonal $= \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

1.3. Singular Value Decomposition (SVD)

<Recall> PSD Matrix A * $\lambda_i \geq 0$

* $\forall v \in \mathbb{R}^n, \; v^TAv \geq 0$

<Def.> $A \in \mathbb{R}^{n\times m}$

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

U: orthogonal $= \begin{bmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{bmatrix} \in \mathbb{R}^{n\times n}$ $\quad$ $\underline{\underline{u_i}} = \frac{1}{\sigma_i} A v_i$ $\quad$ eigenvector of AA^T

left singular vector

V: orthogonal $= \begin{bmatrix} — v_1^T — \\ \vdots \\ — v_m^T — \end{bmatrix} \in \mathbb{R}^{m\times m}$ $\quad$ $A^T v_i = \sigma \underline{\underline{v_i}}$ $\quad$ eigenvector of A^TA

right singular vector

Σ: diagonal $= \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ & 0 & & \end{array}\right] \in \mathbb{R}^{n\times m}$

$\underline{\underline{\sigma_i}} = \sqrt{\lambda_i}$

singular value

Proof. Define $M = A^TA$, PSD, symmetric

Define $\sigma_i = \sqrt{\lambda_i}$, $u_i = \frac{1}{\sigma_i} A v_i$ -- $\{u_i\}$ – orthonormal

$$u_i^T u_j = \frac{1}{\sigma_i}(Av_i)^T \frac{1}{\sigma_j}(Av_j) = \frac{1}{\sigma_i \sigma_j} v_i^T \underbrace{A^T A v_j}_{=\sigma_j^2 v_j} = \frac{\sigma_j}{\sigma_i} v_i^T v_j = \delta_{ij}$$

$$U\Sigma V^T v_i = U \Sigma e_i = U(\sigma_i e_i) = \sigma_i u_i = A v_i$$

$\rightarrow$ $U\Sigma V^T$ and A map the same basis $\{v_i\}$

$\Rightarrow A = U \Sigma V^T$ $\qquad \square$

(Ex.) $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ $\qquad$ $\Sigma = \begin{bmatrix} \sqrt{45} & 0 \\ 0 & \sqrt{5} \end{bmatrix}$

$A^TA = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$ eigenvalues $\lambda_1 = 45$, $\lambda_2 = 5$

eigenvectors $v_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$, $v_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$ $\rightarrow V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$

$u_i = \frac{1}{\sigma_i} A v_i \rightarrow U = \begin{bmatrix} 1/\sqrt{10} & 3/\sqrt{10} \\ 3/\sqrt{10} & -1/\sqrt{10} \end{bmatrix}$

1.4 Norms

<Def.> Vector space E : a norm is a function $\|\cdot\| : E \to \mathbb{R}^+$ that verifies

* Absolute Homogeneity : $\forall \lambda \in \mathbb{R},\ v \in E,\ \|\lambda v\| = |\lambda| \cdot \|v\|$

* Positive Definiteness : if $\|v\| = 0$ for some $v \in E$, $v = 0$

* Triangular Inequality : $\forall v, w \in E,\ \|v + w\| \leq \|v\| + \|w\|$

<Def.> P-norm $(p \geq 1)$ on $\mathbb{R}^n$

$$\|v\|_p = \left[\sum_{i=1}^{n} |v_i|^p \right]^{1/p} \qquad \|v\|_\infty = \sup_{i \in [n]} |v_i|$$

(Ex.) $p=1$. L1-norm : $\|v\| = \sum |v_i|$

$p=2$. Euclidean norm : $\|V\|_2 = v^T v$

<Def.> Frobenius norm on $\mathbb{R}^{m \times n}$ (matrix norm)

$$\|A\|_F = \left(\sum_{i,j} |a_{i,j}|^2 \right)^{1/2}$$

<Def.> Operator norm

$$\|A\|_p = \sup_{v \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

(Ex) Spectral norm $\|A\|_2$

* Remark : $\|A\|_F^2 = \sum_{i=1}^{\min(n,m)} \sigma_i^2(A)$

$$\|A\|_2 = \max_i \sigma_i(A)$$

2. Gaussian Random Variables

<Recall> Gaussian Distribution

$$X \sim N(0,1) \qquad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$Y = \sigma X + \mu \sim N(\mu, \sigma^2) \qquad f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

<Def.> Gaussian Random Vector

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \qquad p(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

μ: expected value

Σ: covariance matrix

$\in \mathbb{R}^{n\times n}$, positive definite, symmetric

$$Y = Ax + b \qquad (A \in \mathbb{R}^{m\times n};\ X, b \in \mathbb{R}^n)$$

$$\left.\begin{aligned} \mathbb{E}[Ax+b] &= A\mu + b \\ \mathrm{Cov}[Ax+b] &= A\Sigma A^T \end{aligned}\right\} \quad Y \sim N(A\mu+b,\ A\Sigma A^T)$$

Gaussian R.V.

(Ex.) $X \sim N(0, I)$: $p(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\|x\|^2}$

---

3 KL-Divergence

<Recall> discrete : $KL(P\|Q) = \sum P(x) \log \frac{P(x)}{Q(x)}$

pdf

continuous: $KL(P\|Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)}\, dx$

KL-divergence of two multivariate Gaussian $\mathcal{N}_1 \sim (\mu_1, \Sigma_1)$, $\mathcal{N}_2 \sim (\mu_2, \Sigma_2)$

trace

$$KL(\mathcal{N}_1 \| \mathcal{N}_2) = \frac{1}{2}\left[\mathrm{Tr}(\Sigma_2^{-1}\Sigma_1) - d + (\mu_2-\mu_1)^T \Sigma_2^{-1} (\mu_2-\mu_1) + \ln \frac{|\Sigma_2|}{|\Sigma_1|} \right]$$

dimensionality