Stat230A Section1 01/27

Wednesday, January 27, 2021

Outline:
1. Eigen-decomposition
2. Projection matrix
3. Vector calculus
4. Multivariate normal
5. Univariate linear statistical estimator
6. Singular value decomposition (SVD)

1. Eigen-decomposition

$P^T P = I_n$
$\det P = \pm 1$

Theorem: If A is an symmetric matrix, then $\exists$ an orthogonal P s.t. $P^T A P = \mathrm{Diag}$.

where λ_i's are eigenvalues of A

column vectors of P are the corresponding eigenvectors.

$$P = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}$$

$$A = P \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} P^T = P\,\mathrm{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)\,P^T$$

$$A^2 = A \cdot A = P\,\mathrm{Diag}(\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2)\,P^T$$

$$(A^2 - A)^2 =$$

$$A = P\,\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)\,P^T = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_n^T & - \end{pmatrix}$$

$$= \sum_{i=1}^{n} \lambda_i v_i v_i^T$$

2. Projection matrix

P is an symmetric matrix that satisfies $P^2 = P$

Claim. The eigenvalues of P is over be either 0 or 1.

Pf. $Pv = P\,\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)\,P^T = \sum_{i=1}^{n} \lambda_i v_i v_i^T$

$P^2 v = P\,\mathrm{Diag}(\lambda_1^2, \ldots, \lambda_n^2)\,P^T = \sum_{i=1}^{n} \lambda_i^2 v_i v_i^T$

$P^2 = P \Rightarrow \sum_{i=1}^{n} \lambda_i v_i v_i^T = \sum_{i=1}^{n} \lambda_i^2 v_i v_i^T$

$\Rightarrow \sum_{i=1}^{n} (\lambda_i^2 - \lambda_i)\, v_i v_i^T = 0$

$\Rightarrow \lambda_i^2 - \lambda_i = 0 \quad \forall i = 1, \ldots, n$

λ_i is either 0 or 1, $\forall i = 1, \ldots, n$ $\square$

Cor. $\mathrm{rank}(P) = \mathrm{trace}(P)$

3. Vector calculus

$f: \mathbb{R}^n \to \mathbb{R}$ (scalar $\to \mathbb{R}^n$)

$$\frac{\partial f(x)}{\partial x} = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}_{n \times 1}$$

Property $\frac{\partial a^T x}{\partial x} = a$ $\quad$ a: any constant vector

$\frac{\partial x^T x}{\partial x} = 2x$ $\quad$ x: any squared matrix

Property $\frac{\partial x^T A x}{\partial x} = 2Ax$ $\quad$ A: any symmetric matrix

For $\mathbb{R}^{n \times n}$ and $x \in \mathbb{R}^n$, we have

$$\frac{\partial x^T x}{\partial x} = 2x \qquad \frac{\partial x^T A x}{\partial x} = 2Ax$$

Pf.

$$b = (b_1, b_2, \ldots, b_n)$$

$$b^T = \begin{pmatrix} b_1^T & - \\ \vdots & \\ b_n^T & - \end{pmatrix} \quad \forall i$$

$$\frac{\partial x^T x}{\partial x} = \begin{pmatrix} \frac{\partial x_1^2}{\partial x_1} \\ \vdots \\ \frac{\partial x_n^2}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 2x_1 \\ \vdots \\ 2x_n \end{pmatrix} = 2x$$

$$\frac{\partial x^T A x}{\partial x} = \begin{pmatrix} \frac{\partial x_1 x_1}{\partial x_1} & \frac{\partial x_1 x_2}{\partial x_2} & \cdots & \frac{\partial x_1 x_n}{\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial x_n x_1}{\partial x_1} & \frac{\partial x_n x_2}{\partial x_2} & \cdots & \frac{\partial x_n x_n}{\partial x_n} \end{pmatrix} = 2Ax$$

4. Multivariate normal distribution

Theorem Assume $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$

1. The marginal distribution of Y_1 is one normal.

$$Y_1 \sim N(\mu_1, \Sigma_{11})$$

$$Y_2 \sim N(\mu_2, \Sigma_{22})$$

2. If $\Sigma_{12} = 0$ then the conditional distribution of $Y_1 | Y_2$ is normal

$$\left(Y_1 | Y_2 = y \right) \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y - \mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)$$

$Y_1 | Y_2$ is independent of the variance $\sim N(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2)$

Example If $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$ then $Y_1 + Y_2$ where

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

Pf.

$$B = (B_1, B_2, \ldots, B_n)$$

$$B^T = \begin{pmatrix} B_1^T & - \\ \vdots & \\ B_n^T & - \end{pmatrix} \quad \forall i$$

$$BY = \begin{pmatrix} B_1^T \\ \vdots \\ B_n^T \end{pmatrix} Y_{n \times 1} = \begin{pmatrix} B_1^T Y \\ \vdots \\ B_n^T Y \end{pmatrix}$$

$$\frac{\partial BY}{\partial Y} = \begin{pmatrix} \frac{\partial B_1^T Y}{\partial Y_1} & \cdots & \frac{\partial B_1^T Y}{\partial Y_n} \\ \vdots & & \vdots \\ \frac{\partial B_n^T Y}{\partial Y_1} & \cdots & \frac{\partial B_n^T Y}{\partial Y_n} \end{pmatrix} = \begin{pmatrix} B_1^T \\ \vdots \\ B_n^T \end{pmatrix} = B$$

Example If $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$ then $Y_1 + Y_2$ where

1. Assume that (Y_1, Y_2) are independent observations.

$$(Y_1, Y_2) \sim N(\mu_1, \sigma^2 I_{2 \times 2})$$

The parameter $\theta = \mu_1$ is the unique solution of $\underbrace{\int_{-\infty}^{\infty} \psi(x)\, dx}_{\text{area of } \int \psi(x) \text{ (positive)}} = 0$

where $\psi(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2}$

The estimator $\hat{\theta} = \hat{\mu}_1$ is the solution of $\underbrace{\sum_{i=1}^{2} \psi(x_i - \theta)}_{\text{area of } \int \psi(x) \text{ (positive)}} = 0$ $\quad$ $\left(\frac{\exp(-x)}{(1 + \exp(-x))^2} \right)$

Under some regularity conditions, $\hat{\theta}$ is consistent for θ and asymptotically normal with

$$\sqrt{2}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2 \Omega^{-1})$$

where $\Omega = \int_{-\infty}^{\infty} \frac{\exp(-x)\exp(-x)}{(1 + \exp(-x))^4}\, dx = \frac{\int_{-\infty}^{\infty} \exp(-x)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$

$$H = \frac{\partial}{\partial \theta} \sum_{i=1}^{2} \psi(x_i - \theta) = \frac{\partial}{\partial \theta} \sum_{i=1}^{2} \psi(x_i - \theta)$$

Maximum likelihood estimator

$(y_1, y_2, \ldots, y_n)$ is a random sample from a distribution with pdf $\frac{\exp(-x)}{(1 + \exp(-x))^2}$

$y_1, y_2, \ldots, y_n \overset{iid}{\sim} N(\theta, \sigma^2)$

where $\theta \in \mathbb{R}^1$ is unknown.

Goal Maximum $\quad$ $\ell(\theta) = \sum_{i=1}^{n} \log \frac{\exp(-x)}{(1 + \exp(-x))^2}$

Maximum likelihood $\quad$ $\ell(\theta) = \sum_{i=1}^{n} \log \frac{\exp(-x)}{(1 + \exp(-x))^2}$

$$\frac{\partial \ell(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log \frac{\exp(-x)}{(1 + \exp(-x))^2}$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \left[\log \exp(-x) - \log(1 + \exp(-x))^2 \right]$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \left[-x - 2\log(1 + \exp(-x)) \right]$$

$$= \sum_{i=1}^{n} \left[1 - \frac{2\exp(-x)}{1 + \exp(-x)} \right]$$

$$= \sum_{i=1}^{n} \left[1 - \frac{2\exp(-x)}{1 + \exp(-x)} \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} \left[\frac{1 - \exp(-x)}{1 + \exp(-x)} \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} \left[\frac{\exp(-x) - 1}{\exp(-x) + 1} \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} \left[\frac{\exp(-x) - 1}{\exp(-x) + 1} \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} \left[\frac{\exp(-x) - 1}{\exp(-x) + 1} \right] = 0$$

$\Rightarrow \hat{\theta}$ is MLE estimator. $\Rightarrow$ $\hat{\theta}$ consistent asymptotically normal

$$\sqrt{I(\theta)}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2 \Omega^{-1})$$

$$\Omega = \int_{-\infty}^{\infty} \frac{\exp(-x)\exp(-x)}{(1 + \exp(-x))^4}\, dx = \frac{\int_{-\infty}^{\infty} \exp(-x)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$$

$H =$ (consistent matrix of $\ell(\theta)$)

$$= \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \frac{\exp(-x)}{(1 + \exp(-x))^2}$$

$$= \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \frac{\exp(-x)}{(1 + \exp(-x))^2}$$

$$\sqrt{I(\theta)} = \sqrt{n}$$

Fisher information

Def Let θ be a scalar

$$\sqrt{I(\theta)}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2 \Omega^{-1})$$

$$\Omega = \int_{-\infty}^{\infty} \frac{\exp(-x)\exp(-x)}{(1 + \exp(-x))^4}\, dx = \frac{\int_{-\infty}^{\infty} \exp(-x)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$$

$$H =