STAT 3690 Lecture 02

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Jan 26th, 2022

Eigendecomposition

- $\mathbf{A}$ is a real square $n \times n$ matrix
- Characteristic equation of $\mathbf{A}$: $\det(\lambda \mathbf{I}_n - \mathbf{A}) = 0$, with identity matrix $\mathbf{I}$
- Eigenvalues of $\mathbf{A}$, say $\lambda_1 \geq \cdots \geq \lambda_n$: n roots of characteristic equation are
- (Right) eigenvector $\boldsymbol{v}_i$: $\mathbf{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$
- Eigendecomposition: $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}$
  - $\mathbf{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ are both $n \times n$ matrices
- Implementation in R: eigen()

Spectral decomposition

- $\mathbf{A}$ is a real symmetric square $n \times n$ matrix
- Then $\mathbf{V}$ is orthogonal, i.e., $\mathbf{V}^\top\mathbf{V} = \mathbf{V}\mathbf{V}^\top = \mathbf{I}$ and $\mathbf{V}^\top = \mathbf{V}^{-1}$
- Spectral decomposition : $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^\top$

Singular value decomposition (SVD)

- Consider a general real $n \times p$ matrix $\mathbf{B}$
- But, obviously, $\mathbf{B}^\top\mathbf{B}$ and $\mathbf{B}\mathbf{B}^\top$ are both symmetric and square
  - They have identical non-zero eigenvalues
  - They are even positive semi-definite, i.e., their eigenvalues are non-nagative
- Then $\mathbf{B}\mathbf{B}^\top = \mathbf{U}_{n\times n}\Gamma_{n\times n}\mathbf{U}_{n\times n}^\top$ and $\mathbf{B}^\top\mathbf{B} = \mathbf{W}_{p\times p}\boldsymbol{\Delta}_{p\times p}\mathbf{W}_{p\times p}^\top$
  - $\mathbf{U}$ and $\mathbf{W}$ are both orthogonal
- SVD:
$$\mathbf{B} = \mathbf{U}_{n\times n}\mathbf{S}_{n\times p}\mathbf{W}_{p\times p}^\top = s_{11}\boldsymbol{u}_1\boldsymbol{w}_1^\top + \cdots + s_{rr}\boldsymbol{u}_r\boldsymbol{w}_r^\top$$
  - Singular values s_{ii} is the ith diagonal entry of $\mathbf{S}_{n\times p}$
  - $s_{11} \geq \cdots \geq s_{rr} > 0$ are square roots of non-zero eigenvalues of $\mathbf{B}^\top\mathbf{B}$ and $\mathbf{B}\mathbf{B}^\top$
  - $\boldsymbol{u}_i$ (resp. $\boldsymbol{w}_i$) is the ith column of $\mathbf{U}_{n\times n}$ (resp. $\mathbf{W}_{p\times p}$)
  - r is the rank of diagonal $\mathbf{S}_{n\times p}$
- Thin/compact SVD
  - Implementation in R: svd()

- Exercise: Is it feasible to apply eigen() only in conducting the thin SVD for a matrix with non-negative singular values (λ_i's)?

"All models are wrong, but some are useful."

— G. E. P. Box. (1976). *Journal of the American Statistical Association*, 71:791–799

Statistical modelling

- What is a statistical model?
  - (Joint) distribution of random variable (RV) of interest
- Rephrase linear regression and logit regression models in terms of distributions

Characterizing/representing univariate distributions

- (scalar-valued) RV X: a real-valued function defined on a sample space Ω
- Cumulative distribution function (cdf): $F_X(x) = \Pr(X \leq x)$
  - right continuous with respect to x
  - non-decreasing w.r.t. x
  - ranging from 0 to 1
- Discrete RV
  - RV X takes countable different values.
  - probability mass function (pmf): $p_X(x) = \Pr(X = x)$
- Continuous RV
  - RV X is continuous iff its cdf F_X is absolutely continuous with respect to x, i.e., $\exists f_X$, s.t.

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\mathrm{d}u \quad \forall x \in \mathbb{R}.$$

  - probability density function (pdf): $f_X(x) = F'_X(x)$.
- Characteristic function
- Moment-generating function

Characterizing/representing joint/multivariate distributions

- Random vector/vector-valued RV
  - $\mathbf{X} = [X_1, \ldots, X_p]^\top$
- Joint cumulative distribution function (joint cdf): $F_{\mathbf{X}}(x_1, \ldots, x_p) = \Pr(X_1 \leq x_1, \ldots, X_p \leq x_p)$
  - right continuous w.r.t. each x_i
  - non-decreasing w.r.t. each x_i
  - ranging from 0 to 1
- Joint distribution of continuous RVs
  - Joint pdf/density:

$$f_{\mathbf{X}}(x_1, \ldots, x_p) = \frac{\partial^p}{\partial x_1 \cdots \partial x_p} F_{\mathbf{X}}(x_1, \ldots, x_p)$$

  - Multivariate normal (MVN) distribution
- Joint distribution of discrete RVs

  - Joint pmf:
$$p_{\mathbf{X}}(x_1, \ldots, x_p) = \Pr(X_1 = x_1, \ldots, X_p = x_p)$$
  - Multinomial distribution
- Multivariate characteristic/moment-generating functions

- Exercise: Suppose that we independently observe an experiment that has p possible outcomes $O_1, \ldots, O_p$ for n times. Let $p_1, \ldots, p_k$ denote probabilities of $O_1, \ldots, O_p$ in each experiment respectively. Let X_i denote the number of times that outcome O_i occurs in the n repetitions. What is the joint pmf of $\mathbf{X} = [X_1, \ldots, X_p]^\top$?

Marginalization

- $\mathbf{X} = [X_1, \ldots, X_p]^\top$ $\mathbf{Y} = [X_1, \ldots, X_q]^\top$, and $q < p$.
- Marginal cdf
$$F_{\mathbf{Y}}(x_1, \ldots, x_q) = \lim_{x_i \to \infty \text{ for all } i > q} F_{\mathbf{X}}(x_1, \ldots, x_p)$$
- Marginal pdf of $\mathbf{Y}$ (when $X_1, \ldots, X_p$ are all continous)
$$f_{\mathbf{Y}}(x_1, \ldots, x_q) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_p) \mathrm{d}x_{q+1} \cdots \mathrm{d}x_p$$
- Marginal pmf of $\mathbf{Y}$ (when $X_1, \ldots, X_p$ are all discrete)
$$p_{\mathbf{Y}}(x_1, \ldots, x_q) = \sum_{x_{q+1}=-\infty}^{\infty} \cdots \sum_{x_p=-\infty}^{\infty} p_{\mathbf{X}}(x_1, \ldots, x_p)$$
- "marginal" is used to distinguish pdf/pmf of $\mathbf{Y}$ from the joint pdf/pmf of $\mathbf{X}$.

Conditioning = joint/marginal

$\mathbf{Y} = [y_1, \ldots, y_q]^\top$ and $\mathbf{X} = [x_1, \ldots, x_p]^\top$

- Conditional pdf of $\mathbf{Y}$ given $\mathbf{X}$
$$f_{\mathbf{Y}|\mathbf{X}}(y_1, \ldots, y_q \mid x_1, \ldots, x_p) = \frac{f_{\mathbf{X},\mathbf{Y}}(x_1, \ldots, x_p, y_1, \ldots, y_q)}{f_{\mathbf{X}}(x_1, \ldots, x_p)}$$
- Conditional pmf of $\mathbf{Y}$ given $\mathbf{X}$
$$p_{\mathbf{Y}|\mathbf{X}}(y_1, \ldots, y_q \mid x_1, \ldots, x_p) = \frac{p_{\mathbf{X},\mathbf{Y}}(x_1, \ldots, x_p, y_1, \ldots, y_q)}{p_{\mathbf{X}}(x_1, \ldots, x_p)}$$

Transformation of random variables (p-dimentional case)

- Let $g = (g_1, \ldots, g_p)$: $\mathbb{R}^p \to \mathbb{R}^p$ be a one-to-one map with inverse $g^{-1} = (g_1^{-1}, \ldots, g_p^{-1})$.
- $\mathbf{Y} = g(\mathbf{X})$ and $\mathbf{X} = g^{-1}(\mathbf{Y})$ are both continuous p-random vectors.
- Jacobian matrix of g^{-1} is $\mathbf{J} = [\partial g_i^{-1}(y_1, \ldots, y_p)/\partial y_j]_{p \times p} = [\partial x_i/\partial y_j]_{p \times p}$.
  - $|\det(\mathbf{J})| = |\det([\partial y_i/\partial x_j]_{p \times p})|^{-1}$ if replace x_j with $g^{-1}(y_1, \ldots, y_p)$
- $f_{\mathbf{X}}$ is known. Then
$$f_{\mathbf{Y}}(y_1, \ldots, y_p) = f_{\mathbf{X}}(h_1^{-1}(y_1, \ldots, y_p), \ldots, h_p^{-1}(y_1, \ldots, y_p))|\det(\mathbf{J})|$$

- Exercise: Let $\mathbf{X} = [X_1, X_2]^\top$ follow the standard bivariate normal, i.e., its pdf is

$$f_{\mathbf{X}}(x_1, x_2) = (2\pi)^{-1} \exp\{-(x_1^2 + x_2^2)/2\}, \quad (x_1, x_2) \in \mathbb{R}^2.$$

  Find out the joint pdf of $\mathbf{Y} = [Y_1, Y_2]^\top$, where $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $0 \leq Y_2 < 2\pi$ is angle from the positive x-axis to the ray from the origin to the point (X_1, X_2), that is, Y is X in polar co-ordinates.