module

M1: Introduction

TMA4268 Statistical learning
07.01.2019

NOTATION:

(multivariate M2)

M3 & M4

n : number of units, observations

Y_i : response or output for unit i : univariate random variable (RV)

Framingham: BP
Spam or not
iris species

- population under study
- draw unit from population
- observe Y_i

simple random sample

continuous (BP) → normal $N(\mu, \sigma^2)$

binary → binomial$(1, p)$

categorical → multinomial

f : pdf/pmf , $F = P(Y_i \leq y_i)$: cdf

probability density/mass function

cumulative distribution function

How to do in R: Rintermediate

$$\mu_i = E(Y_i) = \int y_i f(y_i)\, dy_i \quad \text{or} \quad \sum_{\forall y_i} y_i f(y_i)$$

$$\sigma_i^2 = \text{Var}(Y_i) = E\left((Y_i - E(Y_i))^2\right) = E(Y_i^2) - [E(Y_i)]^2$$

p: number of inputs, predictors, covariates, features

not bf in book

$\rightarrow x_i^T$ = row vector $(x_{i1}\ x_{i2}\ \cdots\ x_{ip})$ ← all variables for obs i, $i = 1, \dots, n$

$$\underset{n\times p}{\mathbf{X}} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdot & \vdots \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_p]$$

bf in book

$$\mathbf{x}_j = \text{column vector } \underbrace{\begin{bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{bmatrix}}_{\text{all } n \text{ observations for variable } j}$$

$j = 1, \dots, p$

Matrix algebra notation: A, B are matrices

A^T transposed

A^{-1} inverse $AA^{-1} = I$ ← identity $\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix}$

AB

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \text{ column vector of 1's}$$

We will need eigenvalues and eigenvectors of A in M6 and M10.

Observe: (Y_i, x_i) $\quad i = 1, \dots, n$ pairs

Y_i: $\mathbb{R}$, $\{0,1\}$, $\{1,2,\dots,k\}$

$x_i \in \mathbb{R}^p$ or some x's $\{0,1\}$ or $\{1,2,\dots,k\}$

often independent pairs

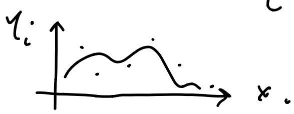
For M2-M9+M11: What is the connection between x_i and Y_i?

BACKGROUND FOR NEXT WEEK:

Let $Y_i = f(x_i) + \varepsilon_i$ $\quad$ where $E(\varepsilon_i) = 0$

$Var(\varepsilon_i) = \sigma^2$

and $\varepsilon_i, \varepsilon_j$ are independent



Then, let $\hat{f}(x_i)$ be an estimator for $f(x_i)$

If $E(\hat{f}(x_i)) = f(x_i)$ then $\hat{f}(x_i)$ is an unbiased estimator for $f(x_i)$

$\left(E(\hat{f}(x_i)) - f(x_i)\right) = \text{bias}$

"NEW" we are not only going to look at unbiased estimators in this course!!

$Var(\hat{f}(x_i)) = E(\hat{f}(x_i)^2) - E(\hat{f}(x_i))^2$

so

$E(\hat{f}(x_i)^2) = Var(\hat{f}(x_i)) + E(\hat{f}(x_i))^2$

The mean squared error is then

$$\underset{\text{mean}}{E}\Big(\big(\underbrace{\underset{\text{estimation}}{\hat f(x_i)} - \underset{\text{true value}}{f(x)}}_{\text{error}}\big)^{2}_{\text{squared}}\Big) = \cdots \text{ next week!}$$

Q: for next week: is this $Var(\hat f(x_i))$?

A: no, $f(x_i)$ need not be $E(\hat f(x_i))$