→ Shivani Tayade & Sravya Bhaskara.

~~Eigenvalues and Eigenvectors~~

Homework I

A positive semidefinite (PSD) matrix is a symmetric matrix s.t. $v^T M v \geq 0$ for all $v \in \mathbb{R}^n$.

1) Prove that a PSD matrix should have all non negative eigen values.

→

$Mx = \lambda x$

$x \neq 0$ eigen vector

λ eigenvalue

$\therefore x^T M x \geq 0$ ..... (~~transforming~~ keeping on one side)

$$x^T . M x = x^T \lambda . x$$
$$= \lambda (x^T . x)$$
$$= \lambda \|x^2\|$$

Since $x^T . x > 0$ as $x \geq 0$ for PSD

$$\lambda . \|x\|^2 \geq 0$$
$$\therefore \lambda \geq 0$$

$\therefore$ all eigen values λ of M are non-negative

2. T/F

'C' correlation matrix for a random vector $X \in \mathbb{R}^n$ can be defined having each entry

$$C_{ij} = \frac{Cov(x_i, x_j)}{\sqrt{Var(x_i) \cdot Var.(x_j)}}$$

a] Can co-relation be used for non-linear relationship

→ False

Co-relation measures linear relationship

Example -

$X = U$ & $Y = U^2$, where $U \sim uniform(-1,1)$

while Y depends non-linearly on X, their co-relation is 0.

b) Does correlation imply causation?

→ No, it ~~does~~ (co-relation) does not imply causation

Example - ice-cream sales and drowning incident may be correlated due to lurking variable - hot-weather - without one being related to another.

c] Is co-relation matrix unaffected by centerⁿ & scaling

→ True.

Correlation standardizes variables (divide by σ) ~~are~~ as it is diffⁿ from mean (centering) and scaling (dividing by a constant) does not affect a matrix co-relation.

Vector Calculus

1.) Let $f(x) = \frac{1}{2} x^T . A . x + b^T . x$.

$A \rightarrow$ symmetric matrix

$b \rightarrow$ vector $\in \mathbb{R}^n$

$\boxed{\nabla f(x)\ ?}$, $\boxed{\nabla^2 f(x) \rightarrow ?}$

→

$\frac{1}{2} x^T . A . x \rightarrow$ differentiates to Ax.

$b^T x \rightarrow$ differentiates to b

$\therefore$ $\nabla f(x) \rightarrow$ gradient $\nabla f(x)$ is $Ax + b$

$\nabla^2 f(x)$ will be A, derivative of Ax w.r.t. x is A (const. in x)

$\therefore \nabla f(x) = Ax + b$

$\nabla^2 f(x) = A$.

2.) Consider the func^c

$$f(u,\lambda) = u^T . X^T . X u - \lambda u^T . u.$$

What is the property of u and λ w.r.t. x at which $\frac{\partial f}{\partial u} = 0$.

$$\frac{\partial f}{\partial u} = 0$$

$$\nabla_u f(u.\lambda) = 2.u. X^T. X - 2\lambda u$$

$\because$ $\partial f / \partial u = 0$

$\therefore$ $u . X^T . X = \lambda u$

λ is a eigen value of $X^T . X$, u is eigen vector.

$\therefore$ Cond^n implies that λ and u must be in that eigen-decomposition relationship

.

Linear Invariance of Regression Coefficients.

$y = \beta_0 + \beta_1 x_i + \epsilon, \quad i = 1, \ldots n$

$x_i \rightarrow$ fixed (non-random)

ϵ_i = uncorrelated, gaussian iid variable with $\mu = 0, \sigma^2$

$$\tilde{x}_i = \frac{x_i - \bar{x}}{S_x}, \quad i = 1 \ldots n$$

$\uparrow$

Standardized x_i, $\bar{x} \Rightarrow$ sample mean

$S_x \rightarrow$ standard sample deviation with n in denominator.

Suppose we fit y_i on $\tilde{x}_i$ by least sq (OLS)

1) What are the least square estimates α_0, α_1 (intercept and slope) for these transformed $\tilde{x}_i$?

$\rightarrow$

$$y_i = \alpha_0 + \alpha_1 \tilde{x}_i + \epsilon_i$$

$\tilde{x}_i \rightarrow$ standardized x

$$\hat{\alpha}_0 = \bar{y} - \hat{\alpha}_1 \cancel{x} \bar{\tilde{x}}$$

$\because$ $\bar{x}_i$ has a mean of 0.

$\therefore \hat{\alpha}_0 = \bar{y}$

$\therefore$ intercept is mean of y

Slope $(\hat{\alpha}_1)$

$$\hat{\alpha}_1 = \frac{\sum(\tilde{x}_i - \bar{\tilde{x}})(y_i - \bar{y})}{\sum(\tilde{x}_i - \bar{\tilde{x}})^2} \quad \text{— } (\beta \text{ book ref}^n)$$

$\bar{\tilde{x}} = 0$

$$\therefore \hat{\alpha}_1 = \frac{\sum \tilde{x}_i (y_i - \bar{y})}{\sum (\tilde{x}_i)^2}$$

$$= \frac{\sum \left(\frac{x_i - \bar{x}_1}{S_x}\right)(y_i - \bar{y})}{\sum \left(\frac{x_i - \bar{x}_1}{S_x}\right)^2} \quad \text{— (1)}$$

$$= \cancel{Cov(x,y)} \cdot \cancel{S_x} = \hat{\beta}_1 S_x$$

$\therefore$ $\hat{\alpha}_1$ is a product of ~~Correlation coeff~~ slope $\hat{\beta}_1$ and standard deviation of x.

$\therefore$ Slope $(\hat{\alpha}_1)$ is scaled by s_x of the original slope

2) Derive the relationship betn the slope estimate and the sample corelation coeffn betn y_i and x_i

→

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - x)^2 \cdot \sum (y_i - \bar{y})^2}}$$

Now, from eqn (1)

$$\hat{\alpha}_1 = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum \cdot \left(\frac{x_i - \bar{x}}{s_x}\right)^2}$$

$$\hat{\alpha}_1 = r_{xy} \cdot \frac{s_y}{s_x}.$$

where. $s_y = \sqrt{\frac{\sum (y_i - \bar{y}_i^2)}{n}}$

detailed derivation

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{— (A)}$$

intercept slope residual.

given OLS estimate.

$$\beta_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} \quad \text{—— (1)}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad \text{—— (2)}$$

Now,

$$\tilde{x}_i = \frac{x_i - \bar{x}}{s_x} \quad \text{—— (a)}$$

$\bar{x} \rightarrow$ mean

s_x = std. deviation $= \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$

Now,

Linear regression

$$y_i = \alpha_0 + \alpha_1 \tilde{x}_i + \varepsilon \quad \text{—— (b)}$$

intercept

slope

Now,

from (a)

$$\tilde{x}_i . s_x + \bar{x} = x_i$$

$\therefore$ from eq^n a, b and A

$$y_i = \beta_0 + \beta_1 \bar{x} + \beta_1 s_x . \tilde{x}_i + \varepsilon_i$$

$$\therefore \alpha_0 = \beta_0 + \beta_1 x$$

$$\alpha_1 = \beta_1 . s_x . \quad \text{— (c)}$$

Now; eq^n (1) and c.

$$\alpha_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} . s_x . \quad \text{— (d)}$$

$$\therefore r_{xy} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 . \Sigma (y_i - \bar{y})^2}}$$

Rearranging eq^n. (d.)

$$\alpha_1 = r.xy . \frac{s_y}{s_x}.$$

where $s_y = \sqrt{\frac{1}{n} . \Sigma (y_i - \bar{y})^2}$

3.) What are the sampling variances of each of these estimates, and their sampling covariance?

$\rightarrow$

$$\because \; y_i = \alpha_0 + \alpha_1 \tilde{x}_i + \varepsilon_i \quad \text{— (1)}$$

$$\alpha_0 = \bar{y} - \alpha_1 \bar{\tilde{x}} \quad \text{— (OLS)}$$

$$\bar{\tilde{x}} = 0 \rightarrow \text{given}$$

$$\therefore \; \alpha_0 = \bar{y}$$

$\therefore$ eq^n (1) $\Rightarrow$

$$y_i - \bar{y} = \alpha_1 \tilde{x}_i + \varepsilon_i$$

$$\therefore \; RSS = \Sigma \left((y_i - \bar{y}) - \alpha_1 \tilde{x}_i \right)^2$$

$$\frac{\partial RSS}{\partial \alpha_1} = -2 \left(\Sigma (y_i - \bar{y}) \cdot \tilde{x}_i \right) + 2\alpha_1 \Sigma \tilde{x}_i^2$$

set $\dfrac{\partial RSS}{\partial \alpha_1} = 0$.

$$\therefore \; \alpha_1 = \frac{\Sigma \tilde{x}_i (y_i - \bar{y})}{\Sigma \tilde{x}_i^2} \quad \text{— (2)}$$

Now,

$$\tilde{x}_i = \frac{x_i - \bar{x}}{s_x}$$

$$\bar{\tilde{x}} = 0 \Rightarrow \tilde{x} = \frac{\Sigma(x_i - \bar{x})}{s_x} = 0$$

Now,

A — $$\Sigma \tilde{x}_i^2 = \Sigma\left(\frac{x_i - \bar{x}}{s_x}\right)^2 = \frac{\Sigma(x_i - \bar{x})^2}{s_x^2}$$

B — Now, $$s_x^2 = \frac{\Sigma(x - \bar{x})}{n^2}$$

from eq^n A & B.

$$\Sigma \tilde{x}_i^2 = n$$

given it is a sample and adjusting for dof

$$\Sigma \tilde{x}_i^2 = n-1 \quad \text{— C}$$

Now,

eq^n 2 becomes

$$\alpha_1 = \frac{\Sigma \tilde{x}_i (y_i - \bar{y})}{\cancel{\Sigma \tilde{x}_i^2} \rightarrow n-1}$$

$$\alpha_1 = \frac{\Sigma \tilde{x}_1 (y_i - \bar{y})}{n-1} \quad \text{— D}$$

(P.T.O)

Now,

$$Var(\alpha_1) = \frac{Var\left(\Sigma \tilde{x}_i (y_i - \bar{y})\right)}{(\Sigma \tilde{x}_i^2)^2}$$

$y_i - \bar{y} \rightarrow$ is the residual. $\rightarrow E(\varepsilon) = \sigma^2$

$$\therefore Var(\alpha_1) = \sigma^2 \cdot \frac{\Sigma \tilde{x}_i}{(\Sigma \tilde{x}_i^2)^2}$$

$$Var(\alpha_1) = \frac{\sigma^2}{\Sigma \tilde{x}_i^2}$$

from. eq^n(c)

$$Var(\alpha_1) = \frac{\sigma^2}{n-1}$$

Intercept

$$\alpha_0 = \cancel{\alpha}\, \bar{y} - \alpha_1 \bar{\tilde{x}}$$

Now,

$$\bar{\tilde{x}} = 0$$

$$\therefore \alpha_0 = \bar{y}$$

$$\therefore Var(\alpha_0) = Var(\bar{y})$$

Now,

$$Var(\bar{y}_i) = \frac{Var(y_i)}{n} = \frac{\sigma^2}{n}$$

$\because y_i$ includes ε_i, variance is equal to residual variance.

3) Covariance of α_0, α_1

$$\alpha_0 = \bar{y} - \alpha_1 \bar{\tilde{x}}$$

$$\text{cov}(\alpha_0, \alpha_1) = \text{cov}(\bar{y} - \alpha_1 \bar{\tilde{x}}, \alpha_1)$$

$\because \bar{\tilde{x}} = 0 \leftarrow$ given

$$\text{cov}(\alpha_0, \alpha_1) = \text{cov}(\bar{y}, \alpha_1)$$

Now,

as we already derived

$$\alpha_1 = \frac{\Sigma \tilde{x}_i (y_i - \bar{y})}{\Sigma \tilde{x}_i^2}$$

$\tilde{x}_i^2$ is zero

This means. $\bar{y}$. ~~do~~ changes in $\bar{y}$ does not affect α

$$\therefore \text{cov}(\bar{y}, \alpha_1) = 0$$

$$\therefore \text{cov}(\alpha_0, \alpha_1) = 0$$

Summary

$$\text{Var}(\alpha_1) = \sigma^2 / n-1$$

$$\text{Var}(\alpha_0) = \sigma^2 / n$$

$$\text{cov}(\alpha_0, \alpha_1) = 0.$$

4) Can you use this estimates to obtain LS estimates for the linear regression model with x_i° & y_i°? How?

→

$$y_i^\circ = \alpha_0 + \alpha_1 \tilde{x}_i^\circ + \varepsilon \quad — ①$$

— Standarized regression model.

Original

$$y_i^\circ = \beta_0 + \beta_1 x_i^\circ + \varepsilon_i^\circ \quad — ②$$

$$\tilde{x}_i^\circ = \frac{x_i - \bar{x}}{s_x}$$

$$\therefore x_i^\circ = s_x \tilde{x}_i + \bar{x}$$

eqn ② ⇒

$$y_i^\circ = \beta_0 + \beta_1 (s_x \tilde{x}_i) + \beta_1 \bar{x} + \varepsilon_i^\circ \quad — ③$$

$$\therefore \alpha_0 = \beta_0 + \beta_1 \bar{x} \quad — ⓐ$$

$$\alpha_1 = \beta_1 s_x \quad — ⓑ$$

Rearranging eqn ⓐ & ⓑ

$$\beta_1 = \frac{\alpha_1}{s_x}$$

$$\beta_0 = \alpha_0 - \beta_1 \bar{x}$$

∴ To convert standardized linear regression (α_0, α_1) to original estimates (β_0, β_1) we $\beta_1 = \alpha_1 / s_x$ and then ~~the~~ compute β_0 with $\beta_0 = \alpha_0 - \beta_1 \bar{x}$

5) What if each x_i was multiplied by 100 before computing $\hat{\alpha}_0$ & $\hat{\alpha}_1$. Would the estimates change! What are the practical implications

→

Original eqn

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Now,

$$x_i' = 100 \cdot x_i$$

$$\therefore x_i = \frac{x_i'}{100}$$

$$\therefore y_i = \beta_0 + \frac{\beta_1}{100} \times x_i' + \varepsilon_0$$

$$\therefore \text{New slope} \Rightarrow \beta_1' = \frac{\beta_1}{100}$$

Original → $\beta_0 = \bar{y} - \beta_1 \bar{x}$

New $\beta_0' = \bar{y} - \beta_1' \bar{x'}$

$$= \bar{y} - \frac{\beta_1}{100} \times (\bar{x}')$$

$$= \bar{y} - \frac{\beta_1}{\cancel{100}} (\cancel{100} \cdot \bar{x})$$

$$\therefore \beta_0' = \beta_0$$

Now,

when we standarize

$$\tilde{x}_i = \frac{x_i - \bar{x}}{s_x}$$

$$x_i' = 100 x_i$$

$$\tilde{x}_i' = \frac{x_i' - \bar{x}'}{s_x'}$$

$$= \frac{100 x_i - 100 \bar{x}}{100 . s_x'}$$

$$\therefore \tilde{x}_i' = \tilde{x}_i$$

$\because$ standarization remained unchanged α_0 and α_1 will also remain unchanged.

Practical implications

1) Slope is inversely related to the scaling factor (100). Therefore as an example, change of scale of m to cm of ~~the~~ x_i will change the slope of 'units of y per m' to 'units of y per cm'.

2) Intercept remains unchanged

3) standarization if done, α_0 & α_1 remains unchanged, as it removes the effect of scaling

6) Provide some examples, where the assumption of independence and corelation of ε_i and x_i is not valid.

→

Classical assumptions of linear regression are
1. Linearity in β_0, β_1, y and x
2. ε_i are independent of each other
3. ~~At~~ ε_i and x_i are not corelated
4. $Var(\varepsilon_i) = 0$ ⇒ homoscedasicity
5. No perfect collinearity
6. $\varepsilon \sim N(0, \sigma^2)$
7. Specialication error.

Examples where independence & corelation violated.

1.) Omitted Variables → ε_i includes the effect of omitted variable that is corelated with x_i. e.g → location variable in house pricing model
2) Measurement error of x_i
e.g. → Income related to self reported education
3) ~~xi~~ When x_i & y_i influence each other
e.g → Modelling supply & price of a given market.

4) Autocorrelation in residual

→ ε_i is not independent & are correlated across x_i

e.g. Time-series data such as stocks, call volume prediction, sales prediction

5) Selection bias.

→ Not random sampling can create a relationship betn x_i and y_i

e.g.:

Studying the impact of one course only for tenured employees.

6) Non-linearity

e.g. Modeling quadratic relation using linear regression model.