Coefficient of determination. (R^2).

$$R^2 = \frac{\text{variation of } Y \text{ explained by the model.}}{\text{Total variation.}}$$

$$= \frac{\text{SS Model.}}{\text{SS Total.}} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

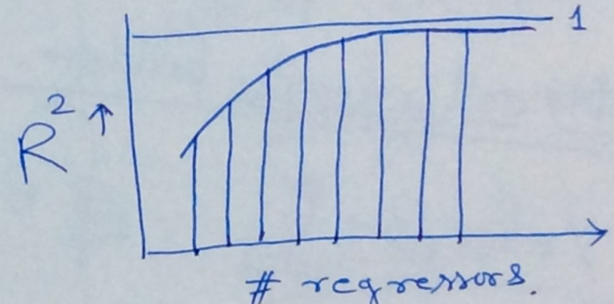
* Larger the value of R^2, the better the model is

* $R^2 \in [0, 1]$

$$SST = \text{SS Model} + \text{SS Error.}$$

$$\Rightarrow \quad 1 = R^2 + \frac{\text{SS Error}}{\text{SST.}}$$

$$\Rightarrow \quad R^2 = 1 - \frac{\text{SS Error}}{\text{SS Total}}.$$

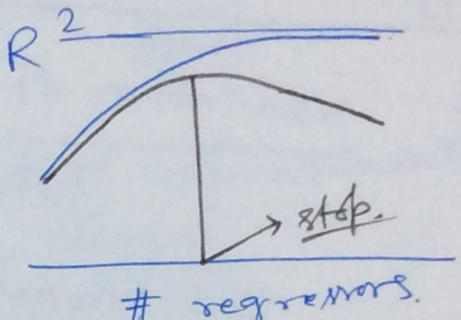


* If we increase the number of regressor variable, the value of R^2 also increase.

Adjusted-R^2 (R^2_{adj}).

$$R^2_{adj} = 1 - \frac{SS\,Error / df(SSE)}{SS\,Total / df(SS\,Total)}$$

$$= 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} < R^2$$



→ Polynomial regression. deside the degree of the model.

Multi collinearity problem.

$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$ $\quad \beta \in \mathbb{R}^{k+1}$

Rank $(X^TX) = k+1$ $\quad$ (illed condition)

Althoorgh $|X^TX| \neq 0$ $\quad |X^TX| \approx 0.$ ⟶ problem.

why? $\quad$ then what?

dimentions are similar in nature (why?)

(Reason.)

① columns are very close to zero.

②. stratified random sampling is not done properly.

Impact:

$$\hat{\underset{\sim}{\beta}} = (X^TX)^{-1} X^T \underset{\sim}{y} = \frac{\mathrm{Adj}(X^TX)}{|X^TX|} X^T \underset{\sim}{y}.$$

As $|X^TX| \approx 0 \Rightarrow Var(\hat{\beta}_j)$ may be unbounded. for some j's

$\Rightarrow |D(\hat{\underset{\sim}{\beta}})|$ or $tr(D(\hat{\underset{\sim}{\beta}}))$ will be too large.

$\Rightarrow \hat{y}_0 = (1\ \underset{\sim}{x}_0)^T \hat{\beta}$ will have large variance.

Consider the expected square error:

$$\hat{\beta} \sim N(\underset{\sim}{\beta}, \sigma^2(X^TX)^{-1})$$

$$E\left[(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})^T(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})\right]$$

↓ max eigen value.

$\lambda_0 > \lambda_1 > \lambda_2 \cdots \lambda_r > \cdots \lambda_k \geq 0$

$$= \sum_{j=0}^{k} E(\hat{\beta}_j - \beta_j)^2$$

nonnegative values arranged in decreasing order. } → product going to zero

$$= \sum_{j=0}^{k} Var(\hat{\beta}_j) = tr(\sigma^2(X^TX)^{-1}) = \sigma^2\left(\sum_{j=0}^{k} \frac{1}{\lambda_j}\right)$$

Note.

(i) X^TX is a symmetric matrix.

(ii) $(X^TX)^{-1}$ is also symmetric.

(iii) Let the eigenvalues of (X^TX) can be arranged as.

$$\lambda_0 > \lambda_1 > \lambda_2 > \lambda_3 \cdots \cdots \geq 0.$$

(iv)
$$\lambda_{min} = \min_{\underset{\sim}{z} \neq \underset{\sim}{0}} \frac{\underset{\sim}{z}^T (X^TX) \underset{\sim}{z}}{\underset{\sim}{z}^T \underset{\sim}{z}} \geq 0$$

symmetric ~~matrix~~ only

$$= \min_{\underset{\sim}{z} \neq \underset{\sim}{0}} \frac{(X\underset{\sim}{z})^T (X\underset{\sim}{z})}{\underset{\sim}{z}^T \underset{\sim}{z}} \geq 0$$

(v) $$\operatorname{tr}(X^TX) = \sum_{j=0}^{k} \lambda_j$$

(vi) $$\operatorname{tr}\left((X^TX)^{-1}\right) = \sum_{j=0}^{k} 1/\lambda_j$$

(vii) $$|X^TX| = \prod_{j=0}^{k} \lambda_j$$

(viii) $$|(X^TX)^{-1}| = \prod_{j=0}^{k} (1/\lambda_j).$$

SVD / PCA

LA

Stat

Spectral representation theorem.

Math

We restrict the use of λ_i such that $\sum_{i=0}^{k} \frac{1}{\lambda_i}$ remains bounded.

Hence we need to focus of the eigen vector corresponding to $\lambda_0 > \lambda_1 \cdots \lambda_p$.

Principal component regression.

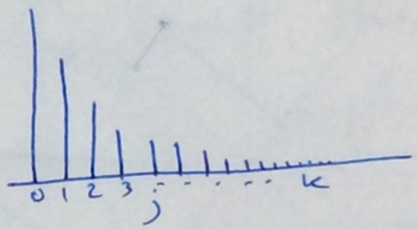
Principal component analysis is dimension reduction technique in multivariate analysis / high dimentional data analysis.

X^TX is a p.s.d matrix / or p.d matrix with $|X^TX| \approx 0$.

We can arrange the eigen values in decreasing order.

$$\lambda_0 > \lambda_1 > \lambda_2 \cdots\cdots > \lambda_k \geq 0$$

$$\begin{array}{ccccc} \downarrow & \downarrow & \downarrow & & \downarrow \\ \underset{\sim}{u}_0 & \underset{\sim}{u}_1 & \underset{\sim}{u}_2 & \cdots\cdots & \underset{\sim}{u}_k. \end{array}$$



$$(X^TX)\underset{\sim}{u}_j = \lambda_j \underset{\sim}{u}_j \qquad j = 0, 1, 2, \cdots k.$$

We can represent all these vectors in a matrix form.

$$P = [u_0 \; u_1 \; \cdots \; u_k]$$

u_j's are orthonormal vectors.

Hence P is an othogonal matrix.

$T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$

there is a matrix A_T

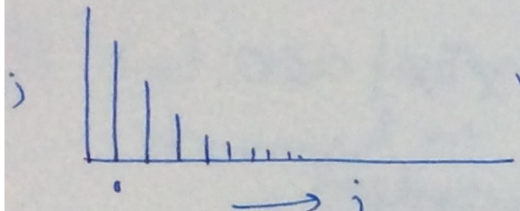
$$T(v) = A_T \underset{\sim}{v}.$$

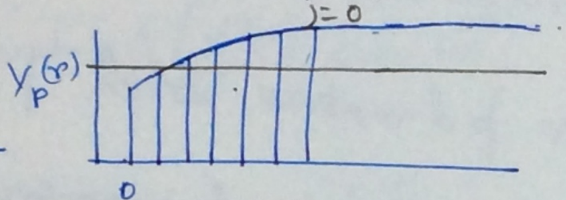
$$sp(\underset{\sim}{v}) = \{c\underset{\sim}{v}\} \qquad c \in \mathbb{R}$$

$$\boxed{A_T \underset{\sim}{v} = \lambda \underset{\sim}{v}}$$

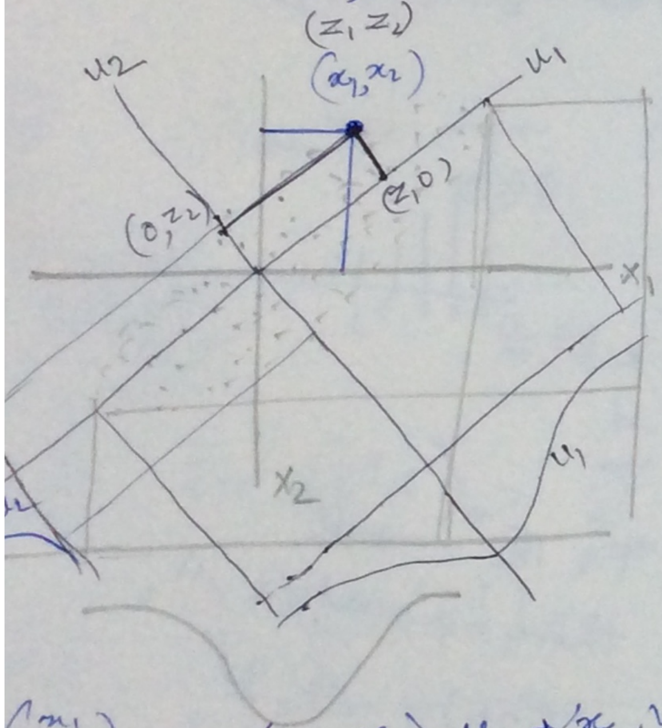
Variation proportion.

$$V_p(r) = \frac{\sum_{i=0}^{r} \lambda_i}{\sum_{j=0}^{k} \lambda_j} \qquad 0 \le r \le k.$$

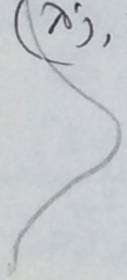




80% or
85% or
90%.



No fixed rule.
to choose how many (λ_j, u_j)s to be considered.



$\{u_j\}$ are new basis elements which are orthonormal.

$\lambda_1 = \lambda_{max} = \max_{x \neq 0} \frac{x^T A x}{x^T x}$ Spectral decomposition theorem.

$\lambda_2 = \max_{\substack{x \neq 0 \\ x \perp u}} \frac{x^T A x}{x^T x}$ wh. $Au_1 = \lambda_{max} u_1$

λ_3 is max in $(\mathrm{sp}\{u_1, u_2\})^{\perp}$

λ_3 is max in $(\mathrm{sp}\{u_1, u_2, u_3\})^{\perp}$.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 \underbrace{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{x_1 \text{ axis}} + x_2 \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{x_2 \text{ axis}}$$

$$= (x_1, x_2)\begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (x_1, x_2)\begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$= \langle \underset{\sim}{x}, e_1 \rangle \cdot e_1 + \langle \underset{\sim}{x}, e_2 \rangle \cdot e_2$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \langle \underset{\sim}{x}, \underset{\sim}{u_1} \rangle \underset{\sim}{u_1} + \langle \underset{\sim}{x}, \underset{\sim}{u_2} \rangle \underset{\sim}{u_2}$$

$$= z_1 \underset{\sim}{u_1} + z_2 \underset{\sim}{u_2}$$

$$\underset{(k+1)}{X^T X} = P D P^T \qquad P = [u_0\ u_1 \cdots u_k] \qquad D = \mathrm{diag}(\lambda_0\ \lambda_1 \cdots \lambda_k)$$

good → (*) dimension reduction. (*) can do prediction in old form of data.

bad → (*) interpretation of regressors are lost. (*) $\underset{\sim}{Y}$ has no role in this dimension reduction.

$$\underset{n \times (k+1)}{Z} = \underset{n \times (k+1)}{X}\ \underset{(k+1)\times(k+1)}{P}.$$

→ new regressors. and we will use $\underline{r}$ $(r+1)$ colums of Z matrix.

$$Y = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}. \qquad \underset{\sim}{\epsilon} \sim N(0, \sigma^2 I_n).$$

$$PP^T = I.$$

$$\Rightarrow Y = XPP^T \underset{\sim}{\beta} + \underset{\sim}{\epsilon}.$$

let $P^T \underset{\sim}{\beta} = \underset{\sim}{\alpha}$.

$$\Rightarrow Y = Z\underset{\sim}{\alpha} + \underset{\sim}{\epsilon}$$

$$\alpha_{(r)} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_r \end{pmatrix} \qquad \underset{\sim}{\alpha} = \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_k \end{pmatrix}$$

$r < k$.

$$\boxed{\underset{\sim}{Y} \approx Z_{(r)} \underset{\sim}{\alpha}_{(r)} + \epsilon.}$$

↓ reduced model.

$$\underline{\hat{\alpha}_{(r)}} = \left(Z_{(r)}^T Z_{(r)}\right)^{-1} Z_{(r)}^T \underset{\sim}{Y}.$$

One of the ways to address multicollinearity $\underline{\text{problem}}$.

we have started with.

$\underset{\sim}{\beta} \in \mathbb{R}^{k+1}$

but $\hat{\alpha}_{(r)} \in \mathbb{R}^{r+1}$

~~Note~~ Now we have to predict for some new value of x which belongs to $\mathbb{R}^{k+1}$

$$\underset{\sim}{\beta} = I\underset{\sim}{\beta}.$$

$$\Rightarrow \underset{\sim}{\beta} = PP^T\underset{\sim}{\beta}.$$

$$\Rightarrow \underset{\sim}{\beta} = P\underset{\sim}{\alpha}$$

$$\Rightarrow \underset{\sim}{\hat{\beta}}_{pc} = P \begin{pmatrix} \alpha(r) \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\hat{y} = \hat{\beta}_{pc}^T \underset{\sim}{x}_{new}$$