Linear Regression

Cp, AIC, BIC, and Adjusted *R*2

Prof. Asim Tewari
IIT Bombay

Regression

$$y = f(x) + \epsilon$$

↑

Find 'f'

$$\hat{y} = h(x)$$

↳ Parametric model/Function $h_{\beta}(x) = \hat{f}(x, \beta)$

E.g. $\hat{f}(x, \beta) = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)}$

ML is to use data to find β.

Minimize Cost Function $L(\beta)$

$$\hat{\beta} = \arg\min_{\beta} L(\beta)$$

Data

$$D = \{(\bar{x}_i, y_i)\}_{i=1}^{n}$$

Null Hypothesis

$\Rightarrow$ $x_1, x_2, \ldots$ are true features or not?

Is β_i 0?

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

Which of the X_i's are actually important ?

Obj: Choose the most imp. 'd' variable from the given 'p' variables.

1.) Subset Selection:

2.) Shrinkage Methods:

3.) Dimension Reduction

Subset Selection

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

$$h_\beta(x, \beta)$$

$X_1, X_2 \cdots\cdots X_p$

0 variable : 1 way $(y = \beta_0)$

1 variable : p ways $(y = \beta_0 + \beta_1 x)$

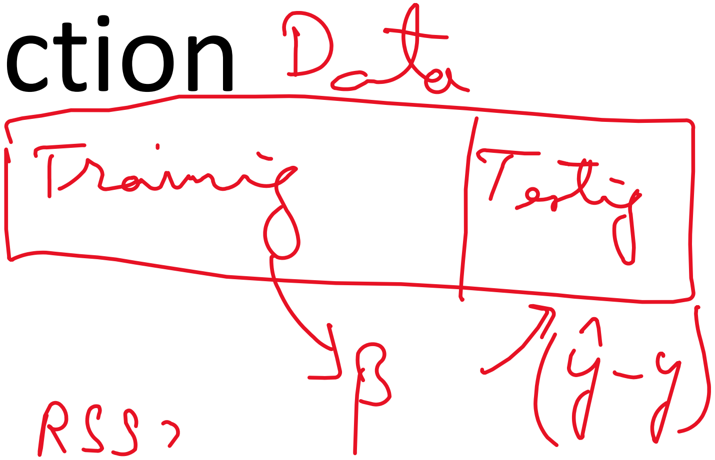
2 Variables : $\frac{p(p-1)}{2} = {}^pC_2$

d variables : ${}^pC_d = \frac{\lfloor p}{\lfloor p-d \; \lfloor d}$

p variables : 1 way

$$1 + {}^pC_1 + {}^pC_2 + \cdots\cdots + {}^pC_d + \cdots + \underbrace{{}^pC_p}_{1} = (1+1)^p = 2^p$$

Subset Selection



Which one is better?

① $\hat{y} = \beta_0 + \beta_1 X_1$

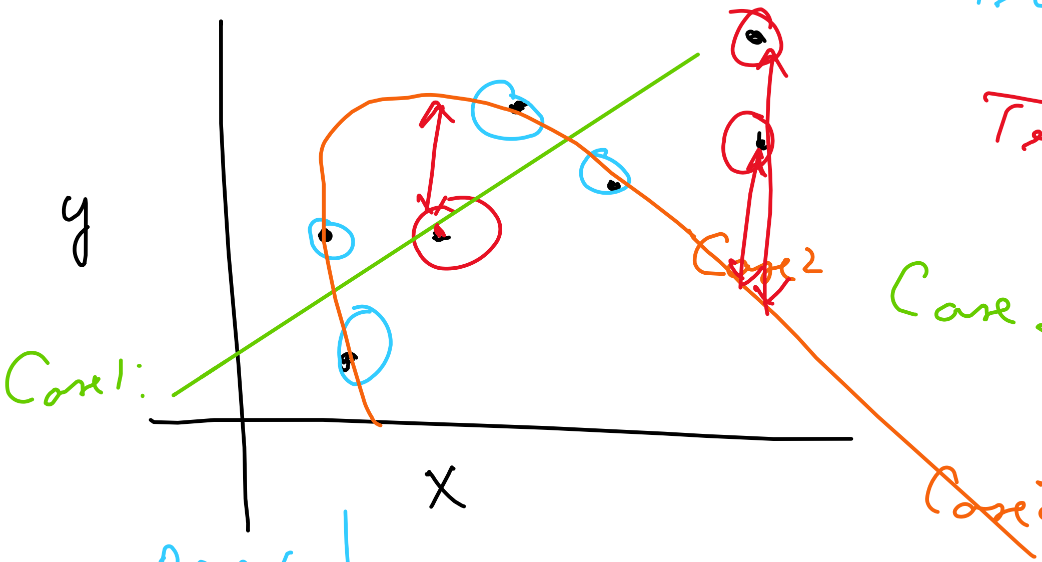
② $\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

RSS?
(Testing?
Training?)

RSS of ① $\geq$ RSS of ②

Better $\Rightarrow$ Lower MSE of Testing

Subset Selection



Train : 4

Test : 3

Case 1: $\hat{y} = \beta_0 + \beta_1 x$

Case2 : $\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$

$\boxed{MSE = RSS/n}$

RSS of case 1 >> RSS of case 2

MSE Training

RSS of Testing of Case 1 < RSS of testing of case 2

MSE Testing

$(MSE)_{Test}$ Case1 < $(MSE)_{Test}$ case2

Cp

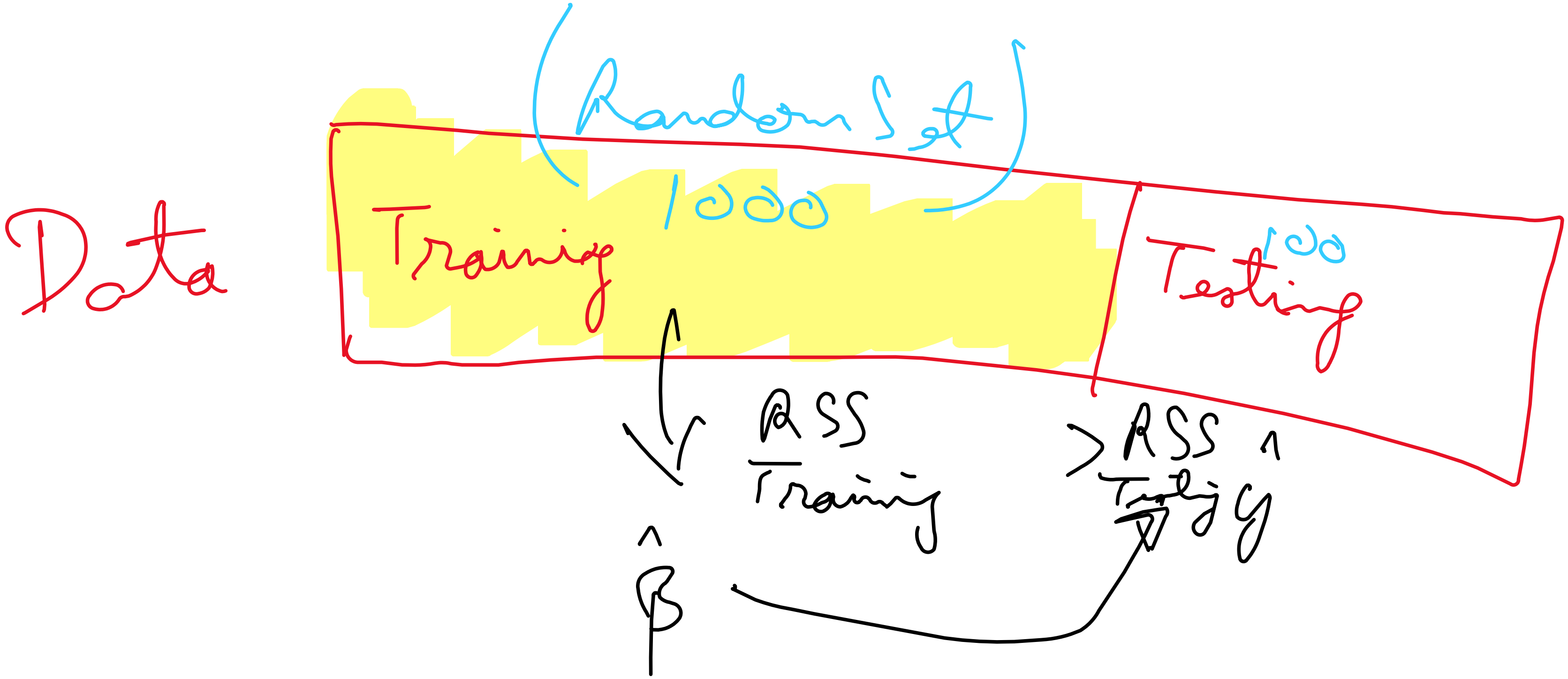
- For a fitted least squares model containing *d* predictors, the *Cp* estimate of test MSE is computed using the equation

$$C_p = \frac{1}{n}\left(\mathrm{RSS} + 2d\hat{\sigma}^2\right)$$

$$MSE = \frac{RSS}{n}$$

- where ˆ*σ*2 is an estimate of the variance of the error associated with each response measurement

Comparison of models with different number of parameters



AIC (Akaike information criterion)

- The AIC criterion is defined for a large class of models fit by maximum likelihood. In the case of the model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

with Gaussian errors, maximum likelihood and least squares are the same thing.

$$\mathrm{AIC} = \frac{1}{n\hat{\sigma}^2}\left(\mathrm{RSS} + 2d\hat{\sigma}^2\right)$$

BIC(Bayesian information criterion)

- BIC is derived from a Bayesian point of view, but ends up looking similar to *Cp* (and AIC) as well. For the least squares model with *d* predictors, the BIC is, up to irrelevant constants, is given by

$\log(n) > 2 \Rightarrow n > 7$

$$\mathrm{BIC} = \frac{1}{n}\left(\mathrm{RSS} + \underline{\log(n)}\, d\hat{\sigma}^2\right)$$

$$C_p = \frac{1}{n}\left(RSS + \underline{2}\, d\, \hat{\sigma}^2\right)$$

Adjusted R^2

$$R^2 = 1 - \frac{RSS}{TSS} \; ; \quad TSS = \Sigma (y_i - \bar{y})^2$$

$$\text{Adjusted } R^2 = 1 - \frac{RSS/(n-d-1)}{TSS/(n-1)}$$

Choosing the Optimal Model

For a fitted least squares model containing d predictors

- *Cp*

$$C_p = \frac{1}{n}\left(\text{RSS} + \underline{2d\hat{\sigma}^2}\right)$$

a measure for MSE for testing

$\frac{RSS}{n}$ = MSE of traing

- *Akaike information criterion* (AIC)

$$\text{AIC} = \frac{1}{n\hat{\sigma}^2}\left(\text{RSS} + 2d\hat{\sigma}^2\right)$$

- *Bayesian information* (BIC)

$$\text{BIC} = \frac{1}{n}\left(\text{RSS} + \log(n)d\underline{\hat{\sigma}}^2\right)$$

- *Adjusted R2*

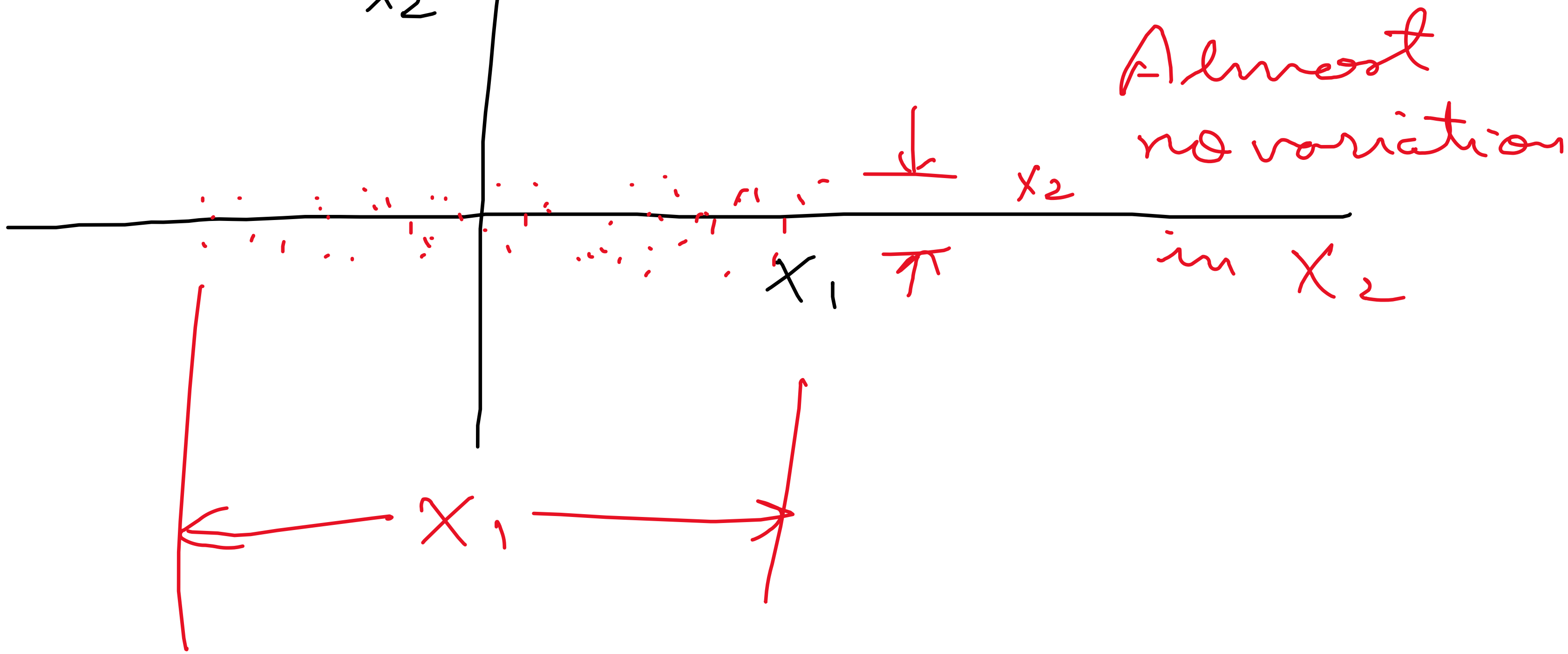
where ˆ σ 2 is an estimate of the variance of the error associated with each response measurement

$$\text{Adjusted } R^2 = 1 - \frac{\text{RSS}/(n-d-1)}{\text{TSS}/(n-1)}$$

Principal Component Analysis

$$y = f(x_1, x_2) \simeq f(x_1)$$

No variation

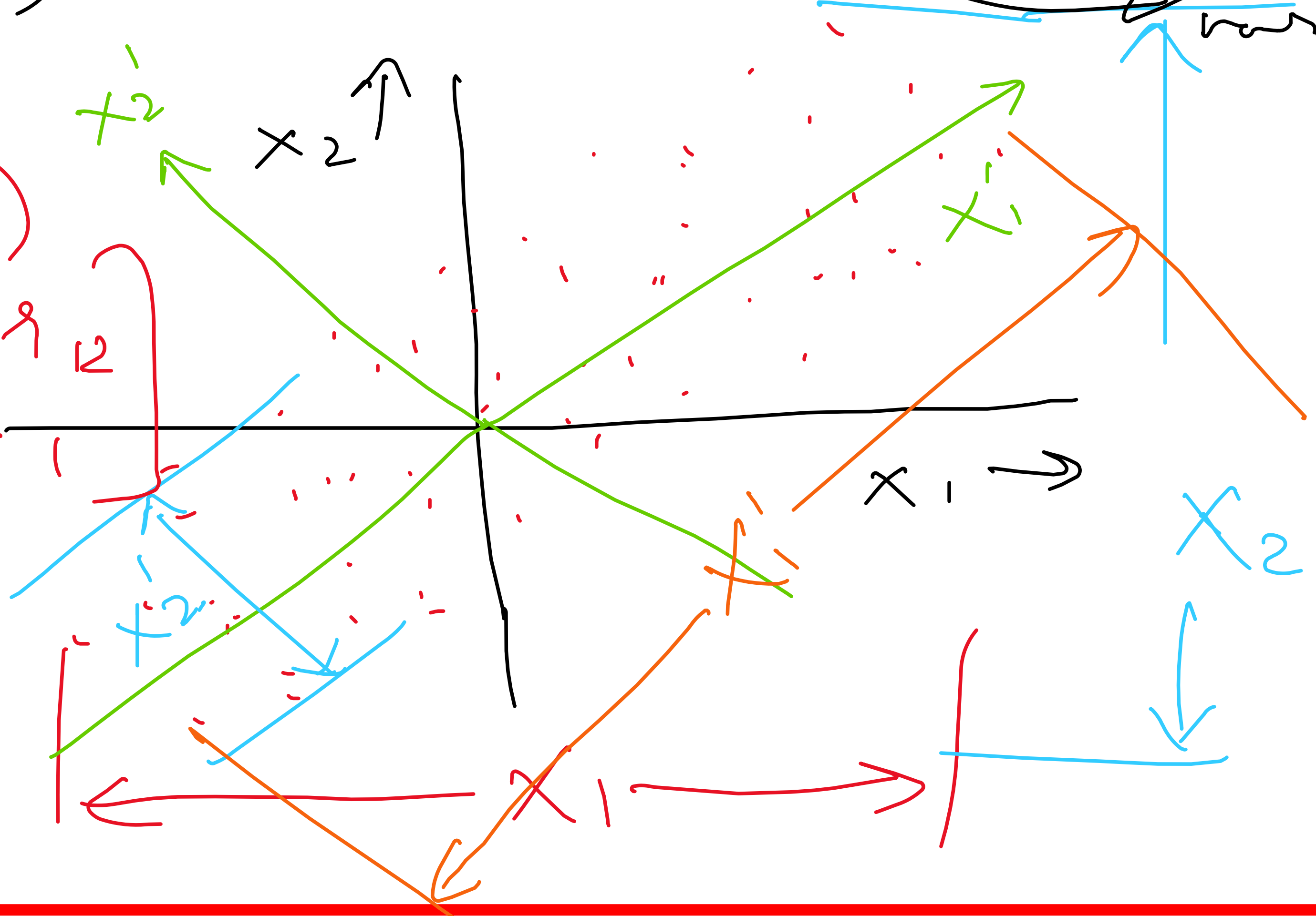


Principal Component Analysis

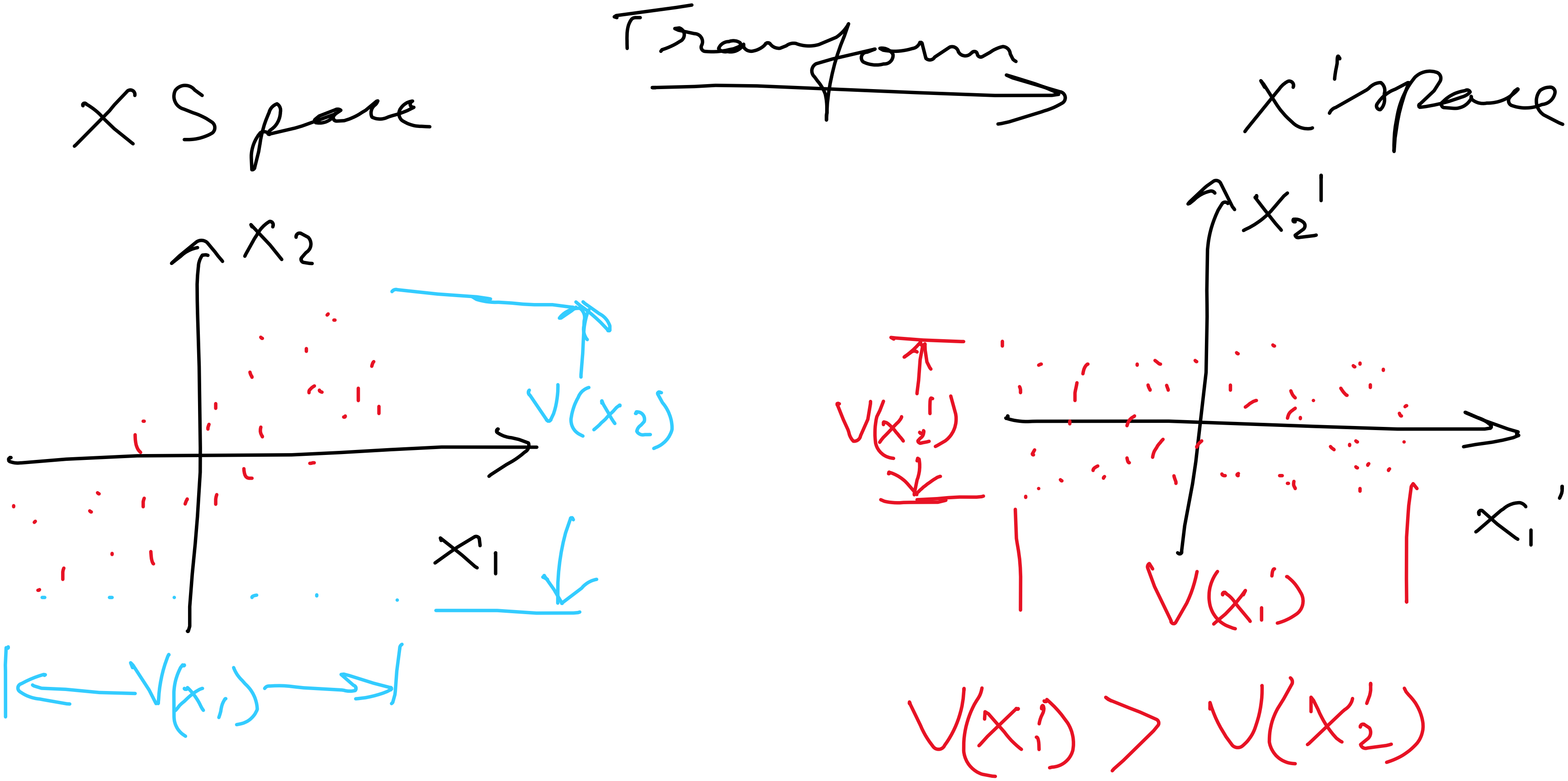
$(X_1, X_2) \xrightarrow{\text{Transform}} (X_1', X_2') \rightarrow$ Low variation

$\text{Corr}(X) = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}$

very close to 1



Principal Component Analysis



$V(x_1') > V(x_2')$

Principal Component Analysis

In 2D space with n data points

X space ⟹ X' space

$$X_{n\times 2} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \\ x_{n1} & x_{n2} \end{bmatrix} \Rightarrow X'_{n\times 2} = \begin{bmatrix} x'_{11} & x'_{12} \\ x'_{21} & x'_{22} \\ \vdots & \\ x'_{n1} & x'_{n2} \end{bmatrix}$$

variance

$$\underset{\substack{\uparrow \\ \text{highest possible variance}}}{V(x_1')} > V(x_2')$$

Principal Component Analysis

In p-dimensional Space

PCA

x space → x' space

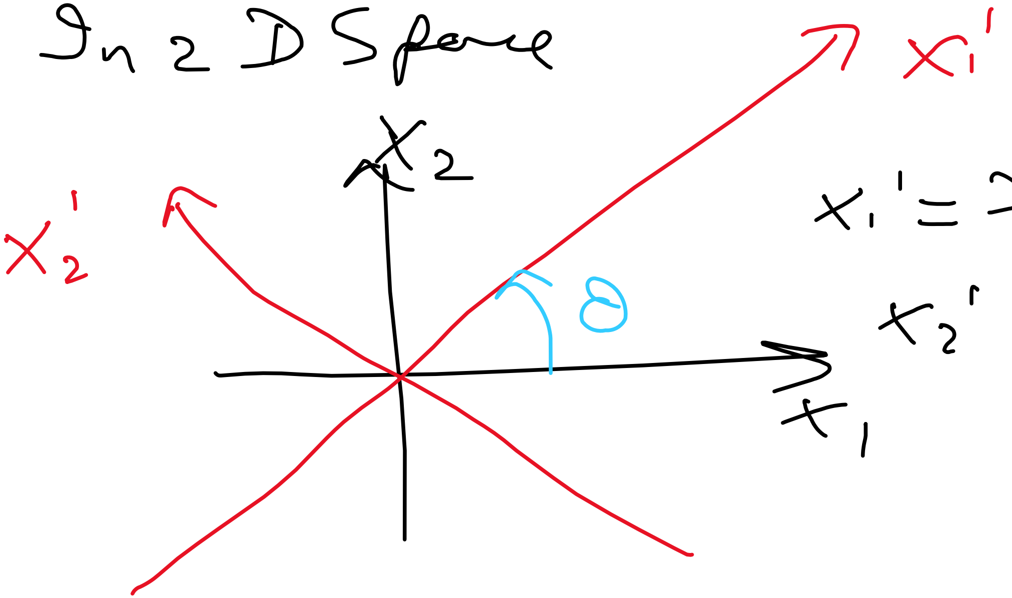
$$X_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$X'_{n\times p} = \begin{bmatrix} x'_{11} & x'_{12} & \cdots & x'_{1p} \\ x'_{21} & x'_{22} & \cdots & x'_{2p} \\ \\ x'_{n1} & x'_{n2} & \cdots & x'_{np} \end{bmatrix}$$

$$V(x_1') \geq V(x_2') \geq \cdots V(x_p')$$

Principal Component Analysis

In 2 D Space



$$x_1' = x_1 \cos\theta + x_2 \sin\theta$$

$$x_2' = -x_1 \sin\theta + x_2 \cos\theta$$

$$\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$X' = A^T X$$

Principal Component Analysis

In 3D $X' = A^T X$

$$\begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \underbrace{\begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$R_3(\theta) = A^T$ For rotation about X_3

Principal Component Analysis

$$R_2(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$R_1(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$$

For a general rotation

$$A^T = R = R_1(\alpha)\ R_2(\beta)\ R_3(\gamma)$$

Principal Component Analysis

In general 3D rotation you can always find an axis that is not changed.

$$\underset{\substack{\uparrow\\ \text{new}\\ \text{axis}}}{X'} = R \underset{\substack{\uparrow\\ \text{old}\\ \text{axis}}}{X}$$

Invariant axis

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\overset{X}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}} = R \overset{X}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}} \Rightarrow (R - I)\underset{\substack{\uparrow\\ \text{Invariant vector}}}{X} = 0$$

Principal Component Analysis

$$\underset{p\times 1}{X'} = \underset{p\times p}{A^T}\ \underset{p\times 1}{X}$$

In 2D

$$A^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

$$\Rightarrow A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = [a_1 \quad a_2]$$

$$a_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} ; \quad a_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

Principal Component Analysis

$$A = \begin{bmatrix} a_1 & a_2 \end{bmatrix}$$

$2\times2 \quad 2\times1 \quad 2\times1$

$$a_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

$$a_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

$$a_1^T a_1 = \begin{bmatrix} \cos\theta & \sin\theta \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \cos^2\theta + \sin^2\theta$$

$$= 1$$

$$a_2^T a_2 = 1$$

$$a_1^T a_2 = \begin{bmatrix} \cos\theta & \sin\theta \end{bmatrix} \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} = 0$$

Principal Component Analysis

$$A^T A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \underline{I} = A A^T = A A^{-1}$$

$$= A^{-1} A$$

$$\Rightarrow A^T = A^{-1}$$

$\therefore$ A is an orthogonal Transfation

Principal Component Analysis

In P-dimension

$$A = [a_1 \;\; a_2 \;\; \cdots \;\; a_p]$$

$$a_i^T a_j = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

In PCA we want this transformation to be such that variance of X' are as follows:

$$V(x_1') \geq V(x_2') \geq V(x_3') \cdots \geq V(x_p')$$

Principal Component Analysis

p - dimensional Space

$$X_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & & & \\ x_{k1} & x_{k2} & \cdots & x_{kp} \\ \vdots & & & \\ x_{n1} & \cdots & \cdots & x_{np} \end{bmatrix}_{n \times p}$$

In set form $X = \{ x_1 , x_2 \cdots x_n \}$

where $x_k = (x_{k1} \; x_{k2} \cdots x_{kp})$

Principal Component Analysis

∴ for the "k^{th}" set

$$x_k = \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kp} \end{bmatrix}^T$$

Similarly

$$; \begin{bmatrix} x'_{k1} \\ x'_{k2} \\ \vdots \\ x'_{kp} \end{bmatrix}^T = \begin{bmatrix} x'_{k1} & x'_{k2} & \cdots & x'_{kp} \end{bmatrix} = x'_k$$

$$\begin{bmatrix} x'_{k1} \\ x'_{k2} \\ \vdots \\ x'_{kp} \end{bmatrix} = A^T \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kp} \end{bmatrix}$$

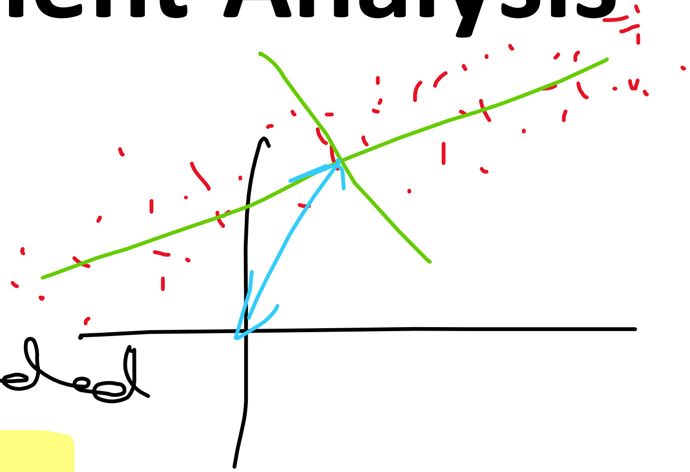
Principal Component Analysis

$$\left[\begin{bmatrix} x'_{k1} \\ x'_{k2} \\ \vdots \\ x'_{kp} \end{bmatrix}\right]^T = \left[A^T \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kp} \end{bmatrix}\right]^T$$

$$\begin{bmatrix} x'_{k1} \\ x'_{k2} \\ \vdots \\ x'_{kp} \end{bmatrix}^T = \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kp} \end{bmatrix}^T A$$

$$\underset{q}{x}'_k = \underset{\uparrow}{x}_k A$$

Principal Component Analysis



$$x'_k = x_k A$$

Translation is also needed

$$\underset{1\times p}{x'_k} = \underset{1\times p}{(x_k - \bar{x})}\,\underset{p\times p}{A}$$

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}\left[\sum X_{k1} \quad \sum X_{k2} \cdots \quad \sum X_{kp}\right]$$

$$\boxed{x_k = x'_k A^T + \bar{x}}$$

Principal Component Analysis

$$x_k = x'_k A^T + \bar{x}$$

variance vector of $x' = \left[\text{var}(x'_1) \quad \text{var}(x'_2) \quad \dots \text{var}(x'_p) \right]$

Covariance of $x' = \underset{\substack{\uparrow \\ p \times p}}{\Sigma_{x'}} = \frac{1}{n-1} \sum_{k=1}^{n} \underset{\substack{\uparrow \\ p \times 1}}{(x'_k)^T} \underset{\substack{\uparrow \\ 1 \times p}}{(x'_k)}$

$$\Sigma_{x'} = \frac{1}{n-1} \sum_{k=1}^{n} (x'_k)^T (x'_k)$$

Principal Component Analysis

$$\nu_{x'} = \frac{1}{n-1}\sum_{k=1}^{n} (x_k')^T (x_k')$$

$$\boxed{x_k' = (x_k - \bar{x}) A}$$

$$\nu_{x'} = \frac{1}{n-1}\sum \left[(x_k - \bar{x})A\right]^T \left[(x_k - \bar{x})A\right]$$

$$= \frac{1}{n-1}\sum_{k=1}^{n} A^T (x_k - \bar{x})^T \cdot (x_k - \bar{x}) A$$

$$= A^T \left[\frac{1}{n-1}\sum_{k=1}^{n} (x_k - \bar{x})^T \cdot (x_k - \bar{x})\right] A$$

$$\nu_{x'} = A^T \underset{\uparrow}{C} A \longleftarrow \text{Covariance matrix of } X$$

Principal Component Analysis

$$v_{x'} = A^T \underset{\uparrow}{C} A$$

Covariance matrix of X

$$C_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} \left[x_{ki} - \frac{1}{n} \sum_{l=1}^{n} x_{li} \right] \left[x_{kj} - \frac{1}{n} \sum_{l=1}^{n} x_{lj} \right]$$

$$i, j \in \{1, 2, \ldots p\}$$

Choose A such that $v_{x'}$ is maximized.

However $A^T A = I$

Principal Component Analysis

Maximise $V_{x'}$ given $A^T A = I$

Constraint optimization.

Construct a Lagrange function

$$L = A^T C A - \lambda (A^T A - I)$$

Sol is A which maximises L

$$\Rightarrow \frac{\partial L}{\partial A} = 0 \Rightarrow CA - \lambda A = 0$$

$$\Rightarrow (C - \lambda I) A = 0$$

Principal Component Analysis

$(C - \lambda I)A = 0$ Solve for A.

This is an Eigenvalue/Eigenvector Problem.

This started by solving DEs:

$$\frac{dy}{dt} = \underset{\substack{\uparrow \\ n \times n}}{A}\, y \quad \Big\}\ n \text{ linear DEs}$$

Sol is of the type $y(t) = e^{\lambda t} X$

Principal Component Analysis

$$y(t) = e^{\lambda t} x$$

$$\Rightarrow \lambda e^{\lambda t} x = A e^{\lambda t} x$$

$$\Rightarrow \lambda x = A x$$

$$\underbrace{(\underset{n\times n}{A} - \lambda \underset{n\times n}{I})\underset{n\times 1}{x}}_{n\times 1} = 0$$

$$\frac{dy}{dt} = A y \Big\} \text{n linear DEs}$$

$$\underbrace{(\underset{p\times p}{C} - \lambda \underset{p\times p}{I})\underset{p\times p}{A}}_{p\times p} = 0$$

$$(A - \lambda I)\underset{\substack{\uparrow \\ n \times 1}}{X} = 0 \quad \Bigg\}$$

It gives eigenvalues of λ and eigenvectors of X

Eg.

$$\left.\begin{aligned} y_1' &= 5y_1 + y_2 \\ y_2' &= 3y_1 + 3y_2 \end{aligned}\right\} \quad \frac{dy}{dt} = \begin{bmatrix} 5 & 1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} y \end{bmatrix}$$

$$\lambda_1 = 6 \; ; \; X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \; ; \; \lambda_2 = 2 \; ; \; X_2 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = c_1 e^{\lambda_1 t} X_1 + c_2 e^{\lambda_2 t} X_2$$

$$y(t) = \underset{\uparrow}{c_1} e^{6t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \underset{\uparrow}{c_2} e^{2t} \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

get from BCs

~~Case 1~~

If we know the eigenvector for $AX = \lambda_1 X$, then we also know the eigenvector for $A^2 X = \lambda_2 X$

$$A^2 X = \lambda_2 X \Rightarrow A(AX) = \lambda_2 X$$

$$\Rightarrow A(\lambda_1 X) = \lambda_2 X$$

$$\Rightarrow \lambda_1^2 X = \lambda_2 X$$

$$\Rightarrow \lambda_2 = \lambda_1^2$$

and they have the same eigenvector

Case 2

$$\text{If } (A + cI)X = \lambda_3 X$$

$$\lambda_3 = (\lambda_1 + c)$$

$$\text{Similarly } A^n X = \lambda_n X$$

$$\text{then } \lambda_n = (\lambda_1)^n$$

$$(C - \lambda I)\underset{p \times p}{\underset{\uparrow}{A}} = 0$$

This is a concatenation of p eigenvectors

$$(A - \lambda I)\underset{n \times 1}{\underset{\uparrow}{X}} = 0$$

$$\underset{p \times p}{\underset{\uparrow}{A}} = (a_1 \; a_2 \cdots a_p)$$

$$(C - \lambda I) A = 0$$

$$A = (a_1, a_2 \ldots\ldots a_p)$$

$$\uparrow \quad \uparrow$$

eigenvectors $\longleftrightarrow$ $(\lambda_1, \lambda_2 \ldots\ldots \lambda_p)$

eigenvalues.

$\Rightarrow$ Variance in X' corresponds to the eigenvalue of $\lambda_1, \lambda_2 \ldots\ldots \lambda_p$

$$CA = \lambda A \Rightarrow A^T C A = \underbrace{A^T \lambda A}_{I}$$

$$\lambda = A^T C A \equiv \mathcal{V}_{x'}$$

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots\ldots \geq \lambda_p$$

Principal Component Analysis

Thus, if we want to capture 95% variance, then we reduce the # of dimensions to q such that

$$\frac{\sum_{i=1}^{q} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \geqslant \frac{95}{100}$$

Principal Component Analysis

Eg. $p = 2, \quad n = 4$

$$X = \{(1,1), (2,1), (2,2), (3,2)\}$$

$$\Rightarrow \quad \bar{X} = \left(2, \frac{3}{2}\right); \quad C = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

The eigenvalues are

$$\lambda_1 = 0.8727, \quad a_1 = \begin{bmatrix} -0.85065 \\ -0.52571 \end{bmatrix}$$

$$\lambda_2 = 0.1273, \quad a_2 = \begin{bmatrix} 0.52571 \\ -0.85065 \end{bmatrix}$$

$$A = [a_1, a_2] = \begin{bmatrix} -0.85065 & 0.52571 \\ -0.52571 & -0.85065 \end{bmatrix}$$

Principal Component Analysis

$$x' = A^T(x - \bar{x}) \Rightarrow x_1' = \begin{bmatrix} -0.85065 & -0.52571 \\ 0.52571 & -0.85065 \end{bmatrix} \begin{bmatrix} 1-2 \\ 1-1.5 \end{bmatrix}$$

$$\Rightarrow \underline{x_1'} = \begin{bmatrix} 1.1135 \\ -0.10039 \end{bmatrix}$$

$$\underline{x_2'} = \begin{bmatrix} 0.2628 \\ 0.42533 \end{bmatrix} ; \quad \underline{x_3'} = \begin{bmatrix} -0.2628 \\ -0.42533 \end{bmatrix}$$

$$x'_{\underline{4}} = \begin{bmatrix} -1.1135 \\ 0.10039 \end{bmatrix}$$

$$X \xrightarrow[A^T]{\text{Transformed}} X'$$

Principal Component Analysis

$X \xrightarrow{A^T} X'$; $A^T = \begin{bmatrix} -0.85065 & -0.52571 \\ 0.52571 & -0.85065 \end{bmatrix}$

X

$\{(1,1), (2,1), (2,2), (3,2)\}$

mean $X = (2, 1.5)$

Covariance of X

$$= \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

X'

$\{(1.1135, -0.10039),$

$(0.26286, 0.42533),$

$(-0.26286, -0.42533)$

$(-1.1135, -0.10039)\}$

Mean $X' = (0, 0)$

$$\text{Cov } X' = \begin{bmatrix} \underline{\overset{\lambda_1}{0.87266}} & 0 \\ 0 & \underline{\overset{\lambda_2}{0.12732}} \end{bmatrix}$$

Principal Component Analysis

$$X \xrightarrow{A^T} X'$$

$$\text{Cov}\, X' = \begin{bmatrix} \lambda_1 & & 0 \\ 0 & \lambda_2 & \\ & & \ddots \; \lambda_p \end{bmatrix}$$

$$X \xleftarrow{A} X'$$

$$X_1, X_2 \xrightarrow{A^T} X_1', X_2'$$

Since x_1' contains 87% of variance, we can drop x_2'

Principal Component Analysis

Let us drop X_2'

Thus we do not take $A = [a_1 \quad a_2]$

but instead $A' = [a_1 \quad 0]$

$$\boxed{X \xrightarrow{(A')^T} X''} \xrightarrow{A'} \sim X$$

$$(A')^T = \begin{bmatrix} -0.85065 & -0.52571 \\ 0 & 0 \end{bmatrix}$$

$$X = \begin{cases} (1,1) \\ (2,1) \\ (2,2) \\ (3,2) \end{cases} \quad X'' = \begin{cases} (1.11351, 0) \\ (0.26286, 0) \\ (-0.26286, 0) \\ (-1.11351, 0 \end{cases} \xrightarrow{A'} \begin{cases} (1.0528, 0.91462 \\ (1.7764, 1.36181) \\ (2.2236, 1.63819) \\ (2.9472, 2.08538) \end{cases}$$

$\sim X$

Principal Component Analysis

$$X = \{(1,1), (2,1), (2,2), (3,2)\}$$

$$\bar{x} = \frac{1}{2} \cdot (4,3)$$

$$C = \frac{1}{3} \cdot \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\lambda_1 = 0.8727$$

$$\lambda_2 = 0.1273$$

$$v_1 = \begin{pmatrix} -0.85065 \\ -0.52573 \end{pmatrix}$$

$$v_2 = \begin{pmatrix} 0.52573 \\ -0.85065 \end{pmatrix}$$

$$E = v_1 = \begin{pmatrix} -0.85065 \\ -0.52573 \end{pmatrix}$$

The projected data

$$Y = \{1.1135, 0.2629, -0.2629, -1.1135\}$$

Inverse PCA yields

$$X' = \{\ (1.0528, 0.91459),\ (1.7764, 1.3618),\ (2.2236, 1.6382),\ (2.9472, 2.0854)\ \} \neq X$$