Curseof D~~i~~mensionality

features

picking only useful features i~~n~~depdent

flier limitations on abrup~~t~~ change in accuracy depending on no of features is called a curse of di~~m~~ensi~~o~~nality

principal compon~~e~~ntalyin

at the no dimension increase itis called as curse Cal theaccuracy is

D

ex

affec~~t~~ed Dame

at ~~77~~ FI

TH

I

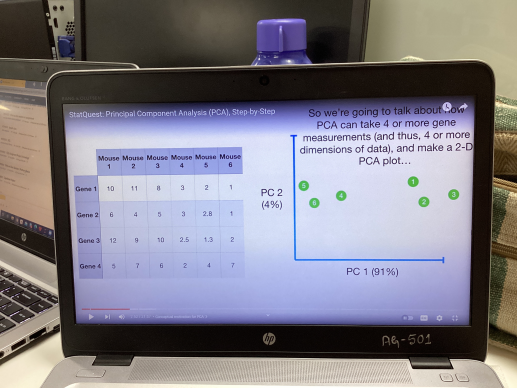
riht lower

OH WE ~~e~~ te

c

ix

g If we measured he genes however we can no longer plo~~t~~ the data he genes require herdimened

So we're going to talk about how MA~~D~~ can take or more gene measurements and thus le or more dimensions of data and make a 2D pc~~t~~ PI 

G

t

a

e

fatly we'll talk abou~~t~~ how Pca can

tell us how accurate the ad graph~~s~~ letsgo back ~~t~~o the data of 2 here aft~~e~~r plotting the

n data as

shown

head

here~~s~~

n

we take the avg

hence measurement for here ~~1~~

here

n

and for gene 2

wai~~t~~

here~~s~~

with the

n avg values we can calculate the

concer~~t~~ a

hewett ~~s~~

cen~~t~~re of the doer

Now we'll shif the data so tha~~t~~ the center is on top o~~f~~

1 ~~ex~~ the origin o o in the graph

now we try to fi~~n~~d a li~~n~~e whi~~c~~h passes through the origin where our centre of data is located

And our Aim is to fi~~n~~d a line in a way that has minimum distance af~~te~~r projection from Deal points

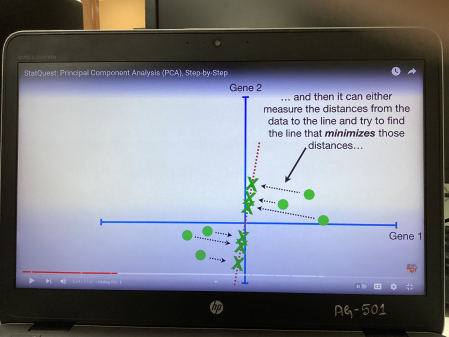
wecan say it li~~k~~e either decrease the distance of real value and projected value

increase the distance f projected points from origin

this the line in

thee care

I ~~t 1~~

how Pat dec~~i~~des thi~~s~~ is the bes~~t~~ fi~~t~~ li~~n~~e

Intenti~~o~~n

Theredistances ge~~t~~ larger when the line fi~~t~~ bet~~t~~er

cb

1 a~~1~~

10 5 ~~72~~

Ty 8~~D~~

t~~h~~en athenbID

now we will take square of distances and sun them I 655 distance cu~~t~~ d~~i~~ct combe ~~te~~

our Aim is it to f~~i~~nd a line

where ther

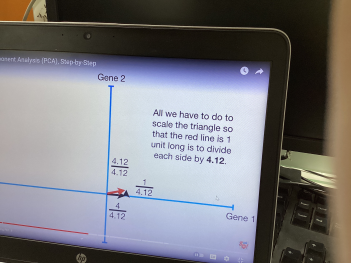
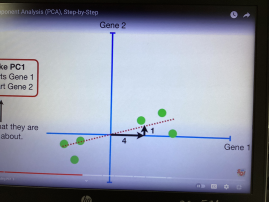
CSS di~~s~~k is larges~~t~~

Th~~i~~s is called pr~~i~~ncipal componen~~ts~~ ~~1~~ for sho~~r~~t

pc~~t~~ is a linear combination of Vari~~a~~bles

In this ez slope of line DoD means 4 parts here ~~I~~

~~la~~nd~~Ip~~a~~rt~~hened

So thee un~~itv~~ector

will guide us and

we can fit the

fittes~~t~~ line

Mt Thi~~s~~ 1 unit long vector consisting of 0.97 parts Lene ~~1~~ and

~~0~~.242 parts here 2 is called

thesingular Vector the

Eigenvector for 10c~~g~~

and the proportions of each gene are called L~~o~~ading scores

SSIdi~~o~~tfor 79 Eigenvalue for Pa sybigenvalueforta Singular Value for Pa

E~~E~~

a ~~te~~r l~~i~~ne to Moaning through fog

thi~~s~~ is our plz

we simply ro~~ta~~t~~e~~ everything 

so that pc is horizontal

then we use the projected

points to find where the samples go in the pa~~d~~ plot

Wakatiperva~~s~~iati~~o~~n for each pc a~~ndree~~ plo~~t~~

Ss distance for Pc Eigenvalue for 191 8s distance for tea E~~i~~genvalue for ME

we convert them into variation around the origin o o by dividing by the sample site minus ~~I~~ ie n D

s~~ldigityd~~ variation for it ssCdiy1rD Variati~~o~~n for 1002

let in the above ex

variati~~o~~n of pc 15

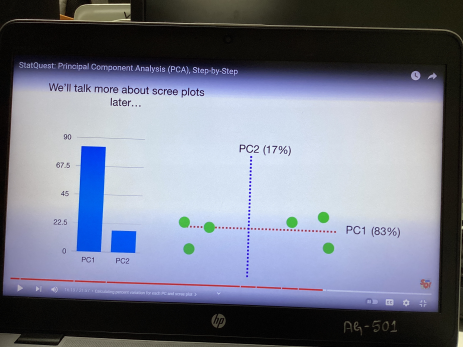
Pg 3 3 74 tree18

9 pc accounts for 15 18 0~~.~~83

the total vari~~a~~tion around the leg

837 I

A 702 3 18 o 17 177 scree plo~~t~~

it is a graph~~i~~cal representation of the percentages of variation that each toaccounts for 

r

it is a line plo~~t~~ of the ei~~g~~envalues of factors a princ~~i~~pal components in an analysis

The scree plo~~t~~ is used to determine the no of factors to retain in an exploratory factor analysis ~~FA~~T con principal components to keep in aped

In~~g~~

Standardize the range of continuous initial var~~i~~ables

compu~~te~~ the covari~~a~~nce matr~~i~~x to iden~~ti~~fy correlations

Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principle Components

Crea~~te~~ a feature vector to dec~~i~~de whi~~c~~h princ~~i~~pal components to keep Recant the data along the

pr~~i~~ncipal Component axe

sina.am fEsiI

mey

ED ED ED

explained as cumulative explained Variance variance

oats

f~~i~~nd

To

pageF~~e~~a~~t~~ureVector Standard~~i~~zedOriginalDorsett

Rought

LinearacminutAnaya

ex we have a cancer drug

It works grea~~t~~ for some people

but not for others

how we decide whom to g~~iv~~e the drug Gene expression can help we dec~~i~~de

lets take ageing and fi~~n~~d

~~9in~~

drug wongL dong notworks

AM~~Y~~oritionth

o~~v~~erlap

~~I~~ J we take scene

cover É

transcripts

Hy if we have 3 D then it w~~i~~ll be more complicated to understand the

distribut~~i~~on

d~~i~~ff btw ~~Ina~~

Pca is useful for plott~~in~~g date wi~~t~~h a lo~~t~~of di~~m~~ensions or lotof geney onto a simple XX thot

However in this core we're no~~t~~ super interested in the genes with the most Var~~i~~ation

Instead we're ~~i~~nterested maximizing the seperatib~~i~~l~~i~~ty b w the two

groups so we can make the bes~~t~~ deci~~s~~ion

CDA is like RA but it fame on maxi muting the Seperatibility among known categories

le~~t~~

I we have a data in 2D and we try to convert it to ID so that we ge~~t~~ the maximum separa~~bilit~~y

if we jut neg~~l~~e~~cte~~d Le~~ep~~ it will create loss of defer

EDDuses both genes to create a new axis and projects thedder

onto this new axis in away to maximise the separation of the two categories

The new axis is created according to two Cr~~i~~teria Simu~~lt~~aneously

Maximi~~z~~e the d~~i~~st min~~i~~mize the btw means variation which LDA calls scatter

oand is represented

by sa within

each cat~~e~~gory

Gimultaneowly T~~h~~oooof

DEFINE ~~if~~

Sq by any dirt

can be big and others

leg le~~t~~ M d

2 heavy but 3 category

f~~i~~st find cute for all points

find diane from each cat cute to main centre

~~YI~~g

we take two lines finally to construc~~t~~ a s~~h~~ear line

D

diff btw LDA and loot

Factor~~A~~naly~~in~~

what is factor analysis

Laten~~t~~ var~~i~~ables

Assumptions in factor analyte purpose of factor analyen

types of fac~~to~~r analysis

Isawe w~~i~~th factor analysis Bane logic of factor analysis

~~of~~

it is a technique that is used to reduce a large no of vari~~a~~bles ~~i~~nto fewer no of factorsit is away of condense the data in many variables into gut few Vari~~a~~bles

it is an example of la~~t~~en~~t~~ variable ~~E~~ model

if you want to go to a hotel

the factors which you consider are

waiting ti~~m~~e

clean li~~n~~en This migh~~t~~ be Staff behav~~i~~our little doff to Taste of food choose any F~~o~~od freehnen ho~~tel~~ f~~o~~od temp~~e~~rature

b~~t~~ Service

FoodQuality

Laten~~t~~Variables

thee are called a laten~~t~~ variables

they are variables that are not directly observed but are inferred from other variables

A mathematical models that aim to explain observed variables in terms of laten~~t~~ variables are called la~~te~~nt variable models

ext moral

happiness in real world

Assumptions in factor analysis

There are no out~~l~~iers in the data Sample s~~ite~~ is supposed to be

greater than the factor

Vari~~a~~bles met be interrelated

Barret fee~~t~~

metr~~i~~c variables are expected

interval data

mult~~i~~variate norm~~a~~lity

to analyze the correlation n~~o~~thing interval of members in particles

Type of factor Analyser

Explorat~~o~~ry factor analytes RFA

used to di~~s~~cover under~~l~~ying structure conf~~i~~rmatory factor analyt~~e~~ C~~F~~A

used to ~~fe~~e~~t~~ if the data fi~~t~~ a priori expectation for data structure

user structural equations modelling ~~E~~FA is div~~ided~~

PCA

common ~~f~~actor Andyin

Image factoring

Maximum li~~k~~elihood Analyeie

Alpha factoring and weigh~~t~~square

Issues with ~~fa~~ctor Analysis

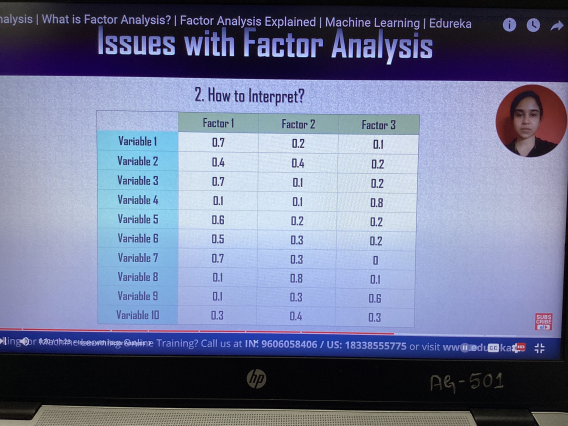
Common

use potat~~o~~ total variance How to i~~n~~terp~~re~~tzvanang How many factors

if you want to find laten~~t~~ variable for many variables then use FA

J you wan~~t~~ to el~~i~~mi~~n~~ate Some variables w~~i~~th hi~~g~~h variance use PC~~A~~

Loading

Factor loading is the correlation coefficient for the var~~i~~able and factor 

to var 9 3 var

variance is devot~~e~~d

Communality hp 1st row

Co 7 7~~1~~0.23~~4~~ Co 112

E~~i~~genvalue for 1s~~t~~ column

Coz~~y~~ co.u~~a~~ 10.25~~4~~

Some times for a part~~i~~cular variable its hows high correlat~~i~~on for more than one factor called as

Crossloading

I

in thi~~s~~ scenario var~~i~~able rot~~a~~tion should be performed

Samplesiz~~e~~

~~G~~ 10variables go observat~~i~~ons

No offactor

j

d~~r~~ab~~l~~y 58,10 faces

make a scree ~~plot~~ Bend in plo~~t~~

lat~~e~~n~~t~~ ~~Rate~~ Eigenvalues ~~a~~

Bas~~i~~c logic of factorAnalysis

Items you want to reduce

Create mathematical combination to fi~~n~~d pc

New comb~~i~~nati~~o~~n from Jeri dual variance and pc

Selec~~t~~ minimal no of factor

Interpret the factory

1

SUD SingularValue Decomposit~~i~~on E~~v~~ery Don matrix factors ~~i~~n~~to~~ Aman Umimtmxn.vn n

is a matrix

U is an orthogonal matrix E is a diagonal matr~~ix~~

UT is again an orthogonal matrix

operative

At a Etat out

~~d~~

t

IA~~T~~.A v~~.fi~~ Da~~t~~a

is same as

Thr a

A A~~T~~ Lieut usta

rerun

Intensit~~y~~

En~~u~~f~~f~~ it

s

z

this is of so

I can be a rec~~ta~~ngle

may as well

manga am

we call them S~~i~~ngular

leftsingular UG

Ven

right singularvector 2

Two geometrical opus in SUD

Rotat~~io~~n U Rot~~a~~t~~i~~on

Stretching E Stream g a r Rotation I s

Coso ~~S~~ino inSVD we call it

S~~i~~no cost unitary transformation

ex

~~6~~ LI~~B~~E~~L~~d ~~É~~

91 People

a~~dethstice~~

text value having

greatest variance

in the table

ad to M~~mff~~

msfmation

will be ache~~i~~ved

from the special matr~~i~~xUED Whosegang a a comb of people

E bi~~g~~gest number Possible

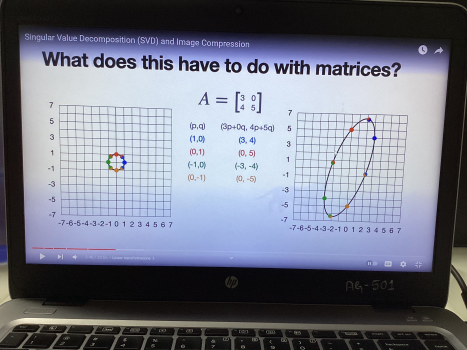
U comb

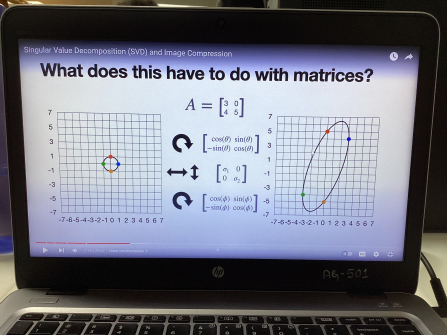
use Li~~n~~ear Algebra

Solve Ax b for non square A Linear regreleion

Gain of PCA

Google F~~a~~cebookNe~~t~~fl~~i~~x etc



y 50

A 3 3 1 2 3

At I I x2

~~a A~~T Iiaxe

a XI P characteristic een to find ev and et

Ma~~ia~~

x 432 0

H a D H X D o

12 7 0 lo k~~o~~

7 12 7 10

these are eigenv~~alue~~

to fi~~nd~~ eigenvectors

xD e~~x~~i ca a

T 1 these are the

~~et~~

we need to perform orthogonali~~s~~ati~~o~~n with the help of

Gram Schmidt orthogonalizati~~o~~n pro cell

hole

pom the Joomla A IET

U and VT need to be orthogonal metri~~c~~s to get them we are performing orogonalisat~~i~~on

i ~~JE~~F~~F~~r

i I~~e~~ R

~~Ar~~e Ya

Ya yaa MTG A~~T~~A V A A~~t~~ u AI~~A~~

CA ~~X~~I X O

oco a 4

4~~4~~,7751 32 0

Si t~~ra~~ceCA to no 2 22 53 de~~t~~ As~~o~~ 1014 26 ~~2~~02

9 522 minor of

diagonals 44 16 60 120 ~~D~~

ye~~s~~ XT sax S3

4 X 2272~~4~~ 1204 0

5 22 120 0

X X O X 12 1203 0 DX x 10 12 4 10 0

ATO 7 12 0

4 0 10212

U O U Us 

row matr~~ix~~

vtCIIJ.am

cos~~t~~ X 12

E I

Ga~~m~~mersald

7.~~5~~

~~Y~~ ~~2~~ ~~7~~

v~~t~~

WE x

2

r~~t~~

we C ~~4 512~~

~~H~~i

rt t~~i~~

nous Gram Schmid proceed we calculate orthogonal matric

VK.E~~E~~~~IFr~~

~~A~~re~~a~~

Ya Ya

YA Ya o

Y

Yo Yoo

cu~~z~~ A 2 3 Ez Ex3

observing and we can see 12,0 are the most repeated values

GEE

higher value

lower value