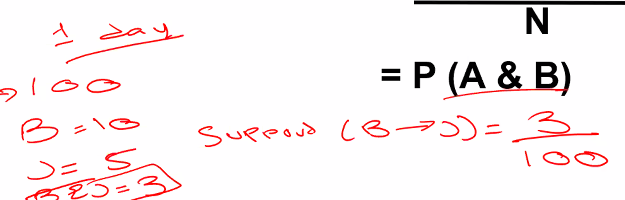
Jam put in front bread is at end.







Lift>1 A and B positive corr

Lift<1 A and B inverse corr.

Appori continues until one item is left.

OBM was missed

1 row 1 transaction.

One hot transaction returns true and false format.

Distinct items made into columns due to one hot encoding.

Min support less more items will show

A->B A antecedents and B Consequents.

Max correlation whole milk.That is why the max support was for whole milk.

We make network graph at end.

Very Important PCA

Exploratory variables reduced through dimensionality reduction

We change the coordinate system.So it can be defined in one dimension.

In third case y values are considered negligible so variation is considered negligible.

In sinosidal variation is too much so no 1-D

PCA makes new origin so that variance is in the middle. So data’s middle point.

Principle axis is made where variance is more. In this case in the direction of X prime other becomes Y prime.

In second there is no dominating axis.But PCA will be implemented.

These two features are combined to create a latent feature.

More data increases model training and execution time.Many of the points are also noise.

The approach where we pick top variables does not work.PCA reduces the variables but keeps the essense of the variables.

Latent features cannot be measured.

Dimensions reduced from 4 to 2.

PCA mostly used in image classification.PCA also used before deciding which algorithm to use it is used to visualize multi dimensional data.It also reduces the PCA.

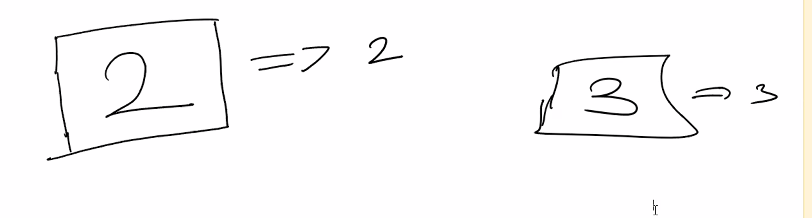
X points are the projections.One dimensional line has the information of both.

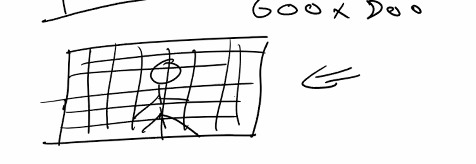
This is PCA(1).

Neighbor 1 PCA(2).

But the information is lost in this.

2% is lost in this.



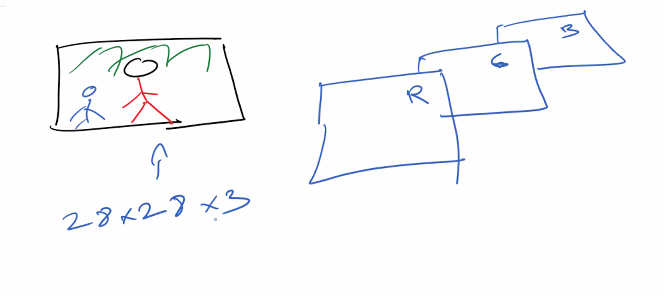


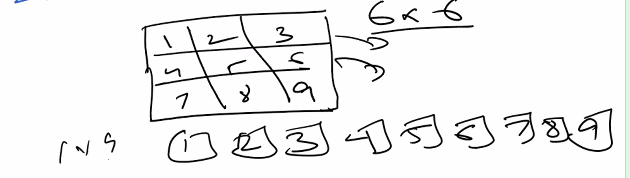
Images cannot be given in 1D

Total pixels=600x800

Color image rgb

Color image=28x28x3

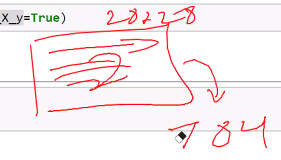




2GB data due to 70000 files.

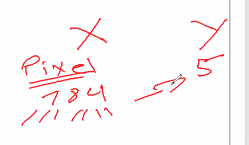
OpenML.org.

Return X\_y returns in 1-D



Greyscale image means 0-255

0=black and 255=white middle is greyish



If we combine the pixels we get the label which is in this case 5.

Imshow is used for pixel by pixel display.

95% information retained and 95% variance covered.

Transform implements mirroring which means the projections.

Support tells how many times 0 occurred similarly for others..

At 100% variance the latent features will be equal to total.

Timer() is used to calculate the time.

Optimal solution for model is at variance 90%.

For deployment we bring from the memory RAM to disk.The model will be saved then.

Pickle is used for this.

Wb means write bit by bit.

Pkl is deployment file.

Total variance can be found by summing all the variance.

We divide the indivial variance by total to calculate the percent for all the values.

For example the first component is covering 9.68.

If we sum all first 87 components it should add up to 90%.

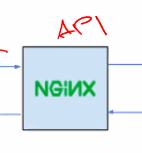
We will make web based application. Any user can import his image and get his prediction.

Open CV is used for importing images.

Rb is retrieve bit by bit.

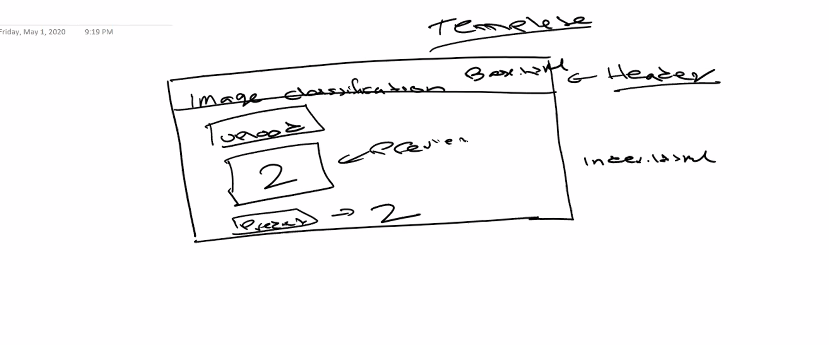
There is no training only prediction in deployment.

Flask is python library that is integrated with web development frameworks like JAVA.



API communicates project to end user.

The entire code is implemented in app.py.



Static is css for styling.

PCA model deploylment not included in test. Sentiment analysis also not included in test.

We will launch using anaconda prompt

Write python in it.

Write cd C:\Users\dell\Downloads\week-8\Flask Deployment\Flask Deployment

Write python app.py

Sentiment analysis is used for textual analysis.

It tells sentiment positive negative or neutral.

We use advanced python libraries for twitter web scraping.

It is a beautiful day (it is a ) stop words. Carry no sentiments.

We save only beautiful day.

Remember twit for web crawling cleansing has different and nlp algorithms are also 2 or more.