**DS, ML, DL, NLP, CV sup topics**

1.Curse of dimensionality reduction.

10 20 30 100 200 500 1000

M1 M2 M3 M4 M4 M5 M5

For certain numbers of features accuracy will can be increased but as the features increases model performance start decreasing as it confused after seeing so many features , which are the features are to focuses on and put importance. And side effect this are overfitting, underfitting, ungeneralised model. And to overcome this we have method like

Forward elimination

Backword Elimation.

Recursive feature elimination.

Feature reduction techniques like PCA……

It refers to the phenomena of strange/weird things happening as we try to analyze the data in high-dimensional spaces.

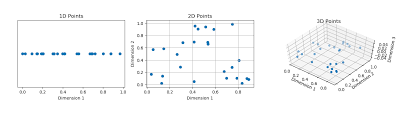
What are the phenomenon:

1.Model will not be generalised.

2.Model can be overfit.

3. Model can be underfit.

4. Model will not be able to use the important features due to the impact of the others features.



Dimension means the features or the independent variable in our data. The main focus in ml is to make a generalise model. Which learn properly from the data. So to have it we need a group of data. Let’s for example 1 one dimensional data will be very dense as we will increase the dimension it the distances among the data will be increased and data will become sparse . And it will so hard to generalise the hence need more data.

But now another point is in high N-dimension data is so close to each to each other that distance of two points from specific one point almost same. That’s why it become harder for algorithm which calculate distances to make the prediction like knn, k-means clustering.

<https://www.analyticsvidhya.com/blog/2021/04/the-curse-of-dimensionality-in-machine-learning/>

https://towardsdatascience.com/curse-of-dimensionality-a-curse-to-machine-learning-c122ee33bfeb#:~:text=Curse%20of%20Dimensionality%20describes%20the,first%20introduced%20by%20Richard%20E.

What is dimensionality reduction and techniques to do it.

2.Eucladian, Manhattan distance.

3.Optimization technique. (Random Search, Grid Search, Baysean optimization.)

4. Differences between the transfer learning and the finetuning.