Day 7 : Challenges in Machine Learning (20 Dec, 24)

#1. Data Collection :

In real world usage, we generally use API or web scraping to fetch the data. And it can be hard as we get diff. formats of data and sometimes it may be not usable.

2. Insufficient Data / Labelled data :

The quantity of data and labelled data can make or break a model. A bad model with huge / good data may outperform a good model with less / bad data.

3. Non-Representative data :

Due to missing data, our data may become biased. This can be a very big issue.

4. Poor Quality data :

If data is bad / not cleaned properly, it can break the model.

5. Irrelevant Features :

Having useless features can make unwanted effect on performance of model and also increase cost of operation.

6. Overfitting :
Feeding too much data which is not diverse. This can show good results on test data but will fail in real world.

7. Underfitting :
Model is very vague and doesn't perform well in any scenario.

8. Software integration :
In the end we want to make the model useful to the audience by implementing it in a software. There are a lot of challenges in doing this as many platforms don't support ML models yet.

9. Offline Learning / Deployment
Discussed previously in 'Day-4'.

10. Cost involved :
Running a ML model and training can be very expensive. Deploying on server is very expensive. Planning to optimize is useful.