

Module 2

Name: Nithin Dsouza

Batch ID: 05062021-9AM(Weekend)

Topic: Machine Learning primer

1. Consider “Correct Bank” as one of the reputed banks in India. It is receiving unlimited calls every day on fraud transactions, the amount has been debited from the users account unknowingly.

Now, the bank is planning to build a model which can provide security to all the accounts before performing any transactions. You as an AI expert must come up with Business Objective and Business constraints.

Solution:

Objective: Minimize Fraudulent and maximize efficiency of transactions.

Constraints: Customer satisfaction and availability of data for all accounts.

2. Find five real time examples of Supervised Learning and Unsupervised Learning

Solution:

Supervised Learning:

1. Cortana or any speech automated system in your mobile phone trains your voice and then starts working based on this training.
2. Based on various features (past record of head-to-head, pitch, toss, player-vs-player) WASP predicts the winning % of both teams.
3. Train your handwriting to OCR system and once trained, it will be able to convert your hand-writing images into text (till some accuracy obviously)
4. Based on some prior knowledge (when its sunny, temperature is higher; when its cloudy, humidity is higher, etc.) weather apps predict the parameters for a given time.
5. Based on past information about spams, filtering out a new incoming email into **Inbox** (normal) or **Junk folder** (Spam)
6. Biometric attendance or ATM etc systems where you train the machine after couple of inputs (of your biometric identity - be it thumb or iris or ear-lobe, etc.), machine can validate your future input and identify you.

Unsupervised Learning:

7. A friend invites you to his party where you meet totally strangers. Now you will classify them using unsupervised learning (no prior knowledge) and this classification can be on

the basis of gender, age group, dressing, educational qualification or whatever way you would like. Why this learning is different from Supervised Learning? Since you didn't use any past/prior knowledge about people and classified them "on-the-go".

8. NASA discovers new heavenly bodies and finds them different from previously known astronomical objects - stars, planets, asteroids, blackholes etc. (i.e., it has no knowledge about these new bodies) and classifies them the way it would like to (distance from Milky way, intensity, gravitational force, red/blue shift or whatever)
 9. Let's suppose you have never seen a Cricket match before and by chance watch a video on internet, now you can classify players on the basis of different criterion: Players wearing same sort of kits are in one class, Players of one style are in one class (batsmen, bowler, fielders), or on the basis of playing hand (RH vs LH) or whatever way you would observe [and classify] it.
 10. We are conducting a survey of 500 questions about predicting the IQ level of students in a college. Since this questionnaire is too big, so after 100 students, administration decides to trim the questionnaire down to fewer questions and for it we use some statistical procedure like PCA to trim it down.
3. Look at the different cases and label them as Underfit and Best fit.

Case 1: Studying for an exam by practicing from the model paper & previous year's paper.

Ans: Best Fit

Case2: Looking at the previous year papers and coming up with important questions and studying only those questions.

Ans: Underfit

Case 3: Preparing for an exam by studying important chapters, previous year's questions and making notes of important points.

Ans: Underfit

4. Let's say you have Real Estate Data. Your data consists of the price of a house, size of the house (square feet), No. of bedrooms, Bulk factor of the house, house location, age or proportion of units built prior to 2000s (or some year), population status, median value of the house, crime rate on the estate etc...

What error function will you use and why?

Ans: Since given feature are Continuous, lets consider our target variable is also Continuous.

So, the best Error function that can be used is MAPE – Mean absolute percentage error. Because it fixed all the mis conception of all other Error functions like Mean error have outliers influence etc.

5. Give 10 real-time examples of Unstructured Data

Ans:

1. Emails.

2. Text files: Word processing, spreadsheets, PDF files.
3. Websites.
4. Social Media.
5. Media (images, video, audio)
6. Mobile data.
7. Communications: live chat, collaboration software.
8. Customer-generated content.
9. Rich media. Media and entertainment data, surveillance data, geo-spatial data, audio, weather data.
10. Internet of Things (IoT). Sensor data, ticker data.