New Lec : 13 Aug 2024 by Abhishek Savaliya

Week-2 remaining stuff

Regulation

↳ do not have linear or inverse linear relationship with innovation

↳ computer science people often thinks like having inverse linear relationship with innovation
- which leads to bad policies

↳ Does any IT company come under MSME? → Midsize or small companies
- no IT company comes under MSME
- the IT companies do not even come under "company"
- they come under "shops"
- hence IT employees are not employees but assistants
- employees can work only 8 hours, assistants ~~are~~ can be ~~made~~ forced to do work even 70 hours a week & it is legal, too.

- There are not enough IT policies in India which leads to exploitation of people

- That's why IT policies are needed in innovation

↳ At some point in 1990's
- there is realization in USA that digital sector have potential to harm us every other sector depends on digital sector
- There is paranoia that very few companies can control most of things in digital era
- Hence we must regulate IT sector rigorously

↳ Google's Ethic council
- every product must pass through this council
- it was fired after ~~schotti~~ ~~Scho~~ stochastic parrot paper
- Google fired entire ethics council after that paper
- companies are fundamentally incapable of self-governance
- bcoz aim of any company is to increase share-holder's values
- companies have job to meet financial aims
- Two ways to make money → either innovate
  → or exploit employees
- they have to exploit workers in order to increase shareholder's value
- since exploitment is need of company, it is weird to ask company to self-regulate
- but in IT sector, companies self-regulate mostly as there are not enough government policies

↳ Ethics Washing

- use of language of ethics to convience society to not regulate the company

(Just like Green Washing → Environment based
Pink washing → female based)

↳ Techno solutionism and reification

- 17th century ~~peo~~ britain example
- personhood & exploitation not solved by making exploitation fast or efficient
- societal problems are not solved by rushing through technology
- You cannot replace social issues which are structural by an accelerant or catalyst
- With AI, you are doing things faster but power is not changed by tech. Tech only increases velocity not power structures.
- Idea that you can solve social issues with tech is called techno solutionism
- Cooking class example for reification

by creating categories, false reality created & everybody are mate to compete in that false reality So everyone works under that delusion and solidifies false reality

- field of ML has reification
  it makes certain categories (like kaggle). Everybody competes under those categories and say it as "state-of-the-art"

- reification is creating realities as creating certain categories

- e.g. money is reification

- we must not confuse reification with scientific objective truth & that's what we do in ML

↳ AI as the scapegoat, or the machine never "understands" or "does" policy

- copystrike on ~~YT~~ Youtube happens by ML algorithm

- like any other ML algorithm, it makes errors & wrongly copystrikes

- who's fault is that?

- platform will say its algorithm's fault

- but the decision to use this algorithm is made by company

- company was financially incentivized bcoz using algorithm saves money

- So people with power are doing decisions. AI is just a tool.

- AI is not neutral tech
- As it is stochastic, it can reproduce past which sometimes we don't want
- Humans are accountable to use this tech

Introduction to ML

↳ What is ML
- Algorithms that improve automatically through experience by using data

↳ Data is unprocessed information
- Information is data after processing

↳ Supervised ML
- training data → annotated data

↳ features
- either ML algo is very sofisticated, you throw raw image and it learns useful things from that image known as "features"
- or your ML algo in not sofisticated & you give features along with image
- features are vectors of n-dimensional data
- for ~~each~~ every feature, you add one dimention in feature vector

↳ while working, ML engine will find out i/p's place in that ~~nt~~ n-dimension space & classify acc. to it

↳ In old times, you have to find good features & give it to model

- It leads to curse of dimensionality
- there will be too much dimensions that you can't computationally train your ML ~~data~~ model
- hence dimentionality reduction done (eg. PCA, SVDC)

↳ Metrics: Correlation, mutual info, class separability

↳ Can you use raw data as features?
- Yes boaz of GPUs
- Google Colab you can use

↳ It's good practise if you ~~use don't or~~ don't use raw data

↳ Jargons
- Cross-Validation
- Your validation data might be stastically just good
- how to robustly check ML algorithm
- hence cross-validation done

eg. divide ~~to~~ data into 10 parts
- every time take 9 parts to train & 1 part to validate
- do it for every ~~possible~~ part
- finally take avg of all validation results

• Active learning
- Suppose model already trained on 80000 images
- now new 20000 images are there
- for many real time ML models, your data keeps on accumulating
- so retraining all data whenever new data comes is not ~~logical~~ good thing to do
- You'll only add those images in training data which are giving you new info
- so you'll first use old ML model to see which new data points are where model is failing
- hence everytime you are retraining, your dataset is increasing by amt of images which are actually adding new info

• Ensemble learning
- You take a bunch of models and join them

• Oracle
- An algorithm which creates synthetic testing data

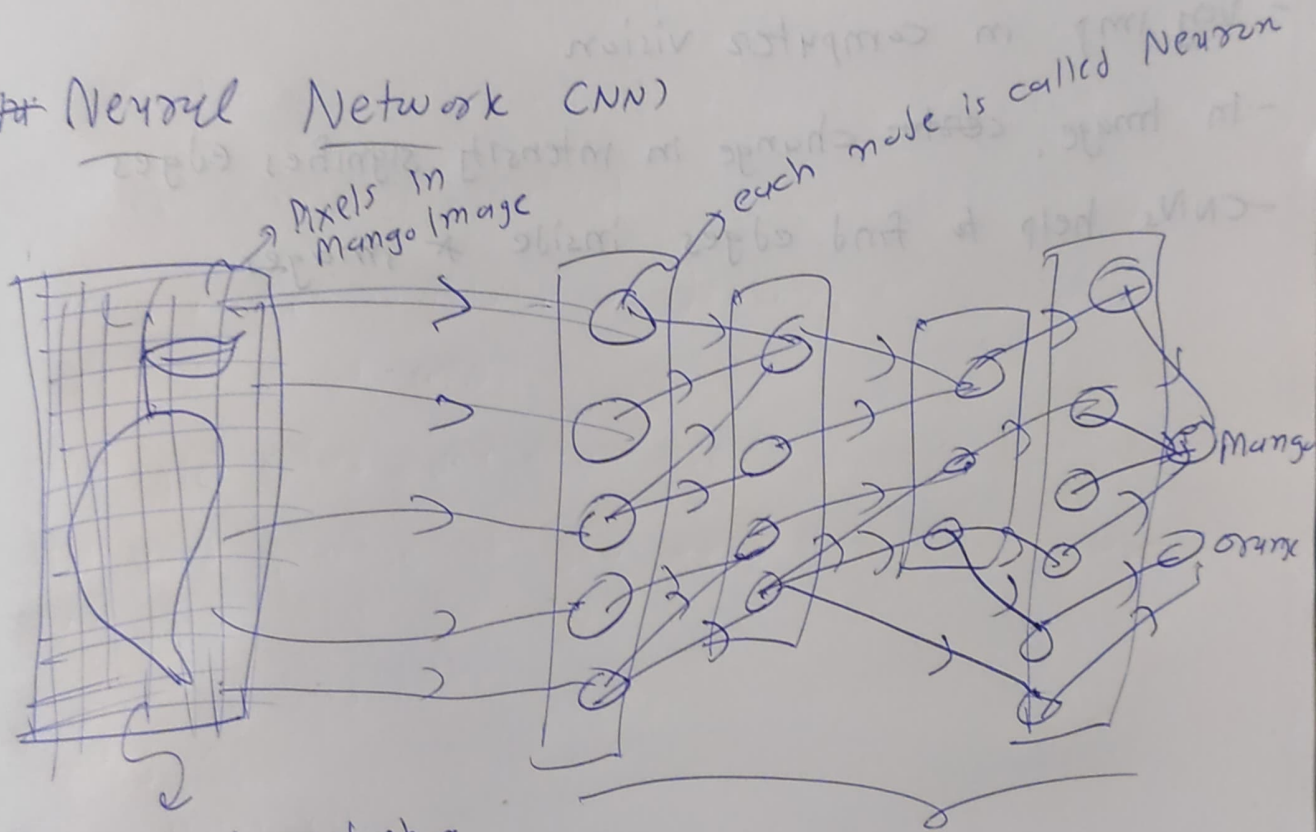
• overfitting
- Your ML algo has become too confident abt training that it starts lacking diversity
- Model won't be alle to look beyond its narrow scope of training
- ~~Its~~ Model which is overfitted is called 'brittel model' bcoz it is easier to break it with anamolous datapoint

- Explainability
- A model which can give explanations why it gave some decision is called explainable

eg. RF is explainable, NNs are not

- Bias
- Model's decision biased towards certain o/ps
- Bias introduced by features

Neural Network (CNN)

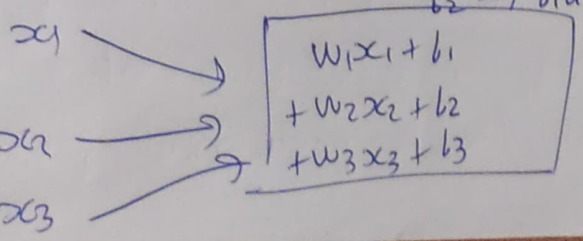


this is not metaphor of a leader of a political party in jail

NN

each Neuron contains weights & biases

Inside a Neuron



answer back propagated to update weights & biases to make model converge to labels

answer should be 0 or 1 or b/w them

↳ A * NN with No intermidiate layer → perceptron

with many intermidiate layers → Deep NN

↳ A NN with layers doing convolutional ~~post~~ processes is called CNN (convolutional Neural Network)

- convolutional process is approximation of gradient or differenciation
- Very imp in computer vision
- In Image, ~~centa~~ change in intensity signifies edges
- CNNs help to find edges inside * images