

Convolution neural networks are used in computer vision.

2 categories of supervised machine learning:

1) In classification labels are discrete.

2) Regression — any kind of ML where the output is a point on a spectrum.

Where do we get the annotation from?

→ Through direct manual work

(inter annotator score,
annotator agreement metrics,
errors are natural in annotations,
disparities make data questionable)

From a policy pov:

Eg. — inner attributes cannot be judged from physical appearance.

Annotator agreement is the 1st requirement of dataset

→ May be available

Determining the best algorithm:-

Not always catching onto the best performing algorithm.

Trade off b/w bias & variance :

Not too much variance but

enough to ~~make~~ not make it brittle.

Basic algorithms in ML

> K-Nearest Neighbour :

Don't have a training set at all.

Computational time: order of 1.

> Decision tree :

Find the feature which divides your data set into 2 parts.

Find the feature which divides your data set closest to the desired output.

Go through the feature in order of their usefulness.

> Bayesian classifier

Simplest one assumes that all the features do not affect each other.

This assumption makes the mathematics simple

Assuming linearity pre-condition.

Sometimes it is not a linear assumption.

$$f(x_1, x_2, \dots, x_n) = L$$

If this is not true...

> Neural networks

Can model any kind of non linear relation.

> Support vector machines

Kernel — will split the data points

↓
Can only do binary problems.

Possible to make a binary classifier into an
n label classifier.

Car, bike, bicycle

Car & others

↓
bike or bicycle.

Multiple ml models in a chain.

> Regression

- Linear

- Logistic



continuous outputs

Unsupervised ML

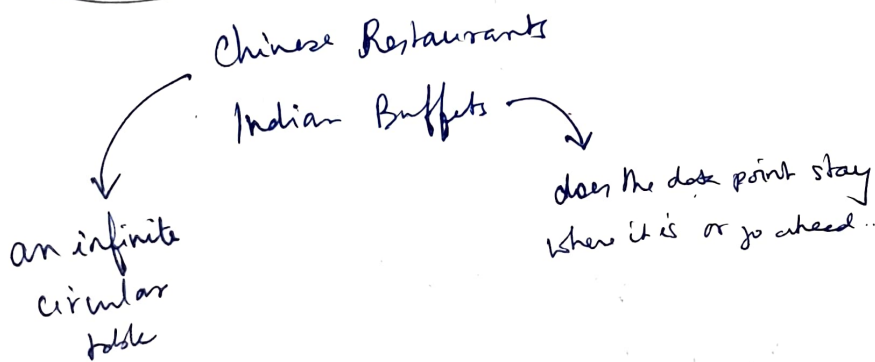
- Not provided w. annotation.
- Try to find internal useful pattern.
- Find out any natural groupings - clustering
Have to give a cluster number

User Guess work needed to prompt -

parametric algorithms.

User inputs - parameters & hyperparameters

Non parametric algorithm



Away to bypass the parameter problem.

Natural language processing → topic model

- ↓
- finds out stat. distribution of all the words.
- find out overlapping themes
- find the ratio of all these themes.

Anomaly detection

Neural Networks in Unsupervised

Generative Adversarial Network (GAN)

Coding

- > Language to communicate w. the computer.
- > Steps of instructions.
- > Languages that operate at the hardware level (machine languages) not used.
- > Human comprehensible instructions. assembled machine language instruction to human understandable language (assembly languages) still used.
- > Languages similar to English
- > Python — high level language
Trade off b/w comprehensibility & resource efficiency

> Compilers and interpreters
↓
↓
take your entire code and translates into machine understandable code
line by line.

> Why use Python?

- Comprehensible
- All code base for machine learning is in Python.

IDLE