Supervised ML:

Classification:

Regression: Output is not discrete, it is a point on a spectrum.

Where do we get annotations from?
1. From humans
2. From natural

Deciding the best algorithm:
- Not only high performance

Bias variance trade-off
- High bias
- High variance

Complexity of algorithm
- The computational cost, order of C.

Classification:
1) K-nearest neighbor - low computational cost for training
majority labeled in k closest points.
2) Decision trees -
Find the feature which divides the dataset.
3) Bayesian classifiers
Assumption is features do not affect each other.
Naïve - is the assumption that features do not affect each other
But it is not true for real world problems. But makes mathematics easy.
4) Neural Networks
To find the non-linear relationships.
Use more than 2 layers.
5) Support Vector Machines.
Kernel will figure out way to split data points.
Kernel is not only linear.
- Can only do binary classification.
Use multiple models to make n-ary classifier.

Regression:
Linear regression
Logistic regression

Unsupervised ML
There are no annotations for the data
Try to find internal.

Clustering
No. of clusters should be provided by users.
Parameters control other parameters - hyper parameter.

Chinese restaurant, Indian buffet → Bayesian non-parametric clustering algorithm

Topic model
Can I find thematic clusters, given the corpus.

Anomaly Detection - find features that don't fit into clusters.

Neural Nets in Unsupervised
GANs

Python:
Language - to communicate with computer.
algorithm - sequence of instructions
Python - High Level Language
Compilers - take entire code and translate into machine
Interpreter - line by line processing

Comments - single line - #
multi-line - ''' '''
Key words and identifiers.
Key words are reserved. They can't use as variables

M T W T F S S

Page No.:

Date:

YOUVA

Output something - print('')

Input - input()

Feature overloading - same instruction is used to use different functions