Supervised ML

Regression : output is point in a spectrum

Classification : output is discrete.

Annotations : manual or Natural sources.
- inter annotation score (multiple annotators).
Annotater agreement doesn't quarentee quality of dataset.

Choosing best algorithm doesn't always depend on accuracy but on computational costs also.

variance : spread of data.

KNN : easiest, no training time

Decision Trees : division based on features.

Bayesian classifier : using probability theory

$\}$ simplest assumes : features don't affect each other
Naive Bayes

- Above all assume linearity pre-condition.
Some times features and labels might not be linearly related.

SVM : binary N.N model based on kernels
kernels split data points (need not be linear)

Dread : Binary classifier.

Unsupervised ML (no annotation)

clustering : guess work of user is needed (no of clusters)
　　　　　　　　　(Parametric algorithms).

Non-parametric clustering ? bayesian theory / statistical
　　　　　　　　　　　　　　　　　　methods.

Topic Models : finds out statistical distribution of
all words, and finds overlapping ones, used in NLP.

Anamaly Detection : finding unusual things,
　　　　　　　　　　useful for surveillance.

Machine languages : Hardware related.

Assembly lang : instructions comprehensible to humans
used in embedding systems.

Python : High level, comprehensible, Inefficient
　　　　　no direct interaction with hardware.

Compiler : translates entire code into machine code.

Interpreter : translates line by line.

feature overloading : multiple functions depending
　　　　　　　　　on context.