Artificial Intelligence

- thinking likes Human
- acting likes Human → Turing Test → in many domains computer is better (Math)
- thinking Rationally
- acting Rationally

the best one

① Raw Data Processing: Images...
Deep learning is better

② if you want to know why we get that and you can live with lower accuracy machine learning is better

Areas of AI

① Computer vesion (CV)

② natural Language Procecing

③ Robotics (moving)

④ Knowledge representation (storing) and deductioning (logic)

⑤ machine Learning (It improves others areas)
- a) Task T
- b) Pernance P
- c) Experince E (Data)

i) Deep learning the best in some tasks

Machine Learning

students score have relation ships → So we use regression

need enough sample for every class

in Deep learning these will be learned automatically from the raw data

i) supervised ML
- a) classification (output: class) — y is discrete → binary class / multi class
- b) regression (output: number) — y is continuous → (Time series) stock market

Data = Independing variable (features / predecter) X + target Label y

f: try to learn the relationship (the program) approximatly

Data without Lables

ii) unsupervised ML (finding structure in data)

we still have Data but we Learn by
1) clustering  grouping data
2) Dimensionality Reduction Project too feacures to just 3 → not from too

70% 30%

(iii) Reinforcment learning

- state
- action
- environment
- reward

train + test should be different to avoid memorizing

we want learning

Properties of Sets → iid
- identical destribution
- independent (no seq or pattren)

- good target of ML ⇒ Human Level
- minimum target ⇒ random Level

Nearest neighbour (1967)

(NON-parametric)

→ Similarity

→ Distance → Ecludian $\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$

we use validation set to select K value. because if we depend on test set memorize may affect the result

we use (K) nearest to avoid overfitting (just good for training set) if K=1

hyperparameter

underfitting if K=n

find it by:- ① train-validation set
② cross-validation

Data → train / test / validation

bigger data-set → bigger training set

* you have to make sure all features have the same unit to Avoid different affecting while using distance (one feature dominate the result)

bye-
1- Standered scaling $\frac{x-u}{\sigma}$
2- divid by max [0,1] (MAximum Abslote)
3- divid by Range $\frac{x-min}{max-min}$ (min-max scale)

When data set is Small we use V-fold. u

cross-validation:-

1	2	3	4

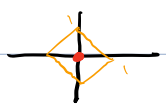
	train	test
exper 1	1,2,3	4
	.	3
	.	2
	i	1

* increase V-fold if data is small
* more features → more performance → relavent

Leave one out ⇒ just one sample for testing

* Lp distance:-

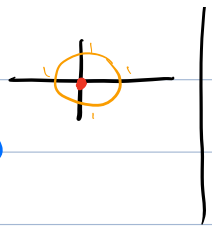
① L_1 = Manhatten $|x_1-x_2|+|y_1-y_2|$



increase K ⇓ underfitting

② L_2 = Eclidian $\sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}$

③ L_∞ = $\max(|x_1-x_2|, |y_1-y_2|)$



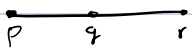
* measurs to be metric: should be follow:

① $d_{p,p} = 0$

② $d_{p,q} = d_{p,q}$

③ $d_{p,q} \neq 0$

④ $d_{p,q} + d_{q,r} \geq d_{p,r}$



} when you replace the measurment to be sure it will work

* Improve KNN:- (faster in test time) ← liner

split on features
① KD-tree (log) → works with small number of features you can't Prun if there are lots

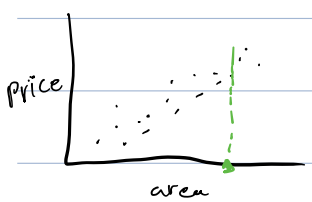
② Ball tree → faster in large number of features (50 ... 200)
↘ higher? basic KNN

* Error Analysis → in validation → we don't do any EA in testing

* if errors are small number → Leave it as it's

* we use confusion matrix after validation to evaluate our validation

nn Regression



choose the price of nearst without

just classifying