Week - I.

1. Define of AI.

Artifical intelligence is a field, which combines computer science and robust datasets, to enable problem-solving. concerned with the development of algorithms that allow computers to learn without being explicitly programmed.

2. AI的一些应用.

Automation, Machine Learning, Computer ~~Vision~~ Vision, Natural Language Processing,

Self-Driving Cars

3. Probability theory

Joint Probability Distribution (JPD) → conditional probabilities → Bayes Network

hard to creat.
very expensive, impossible to store because of their size.

reduce the size of JPD (conditional probability table)
↑
conditional independence

CPT.

4. Meachin Learning is a branch of AI, which focuses on methods that learn from data and make predictions on unseen data.

5. Learning is a goal-directed process of a system that improves the knowledge or the knowledge representation of the system by exploring experience and/or prior knowledge.

对学习机(因素) { 1. which component is to be Improved.
2. prior knowledge
3. representation of data.
4. feedback.

6. Decion Trees., classifier.

A simple yet effective form of Learning from examples.
map objects with a certain set of discrete attributes to discrete values.
based on the value of those attributes.

7. Entropy is a measure of disorder or uncertainty.

Information gain: measure the reduction in entropy or surprise by splitting a dataset according to a given value of a random variable.

limitations: Noise, Overfitting, Missing data, Multi-valued attributes
Continuous-valued attributes.

8. state and explain the three ways that an experiment can fail when uncertainties are not considered appropriately.

① LAZINESS (laziness): too much work to list the complete set of antecedents (前因后果) or consequents needed to ensure an exceptionless rule and too hard to use rules.

② THEORETICAL ignorance (theoretical): Medical science has no complete theory for the domain.

③ practical ignorance: Even if we know all the rules, we might be uncertain about a particular patient because not all the ~~necessary~~ necessary tests have been or can be run.

9. THREE main types of learning in AI.

The type of learning depends on the feedback to learn ~~for~~ from, there are ~~3~~ many types of learning, namely unsupervised, supervised, reinforcement learning.

① Unsupervised learning: the agent learns patterns in the input even though no explicit feedback is supplied.

(Ex. clustering, Scenario ?, Dimensionality reduction).

② Supervised learning: the agent observes some example input-output pairs and learns a function the maps from input to output/classification

(Ex. NNs, Numerical classifier functions, SVM, repression).

③ Reinforcement learning: the agent learns from a series of reinforcement-rewards or punishments.

(Ex. learning to play GO?, Game).

Week-2.

DL

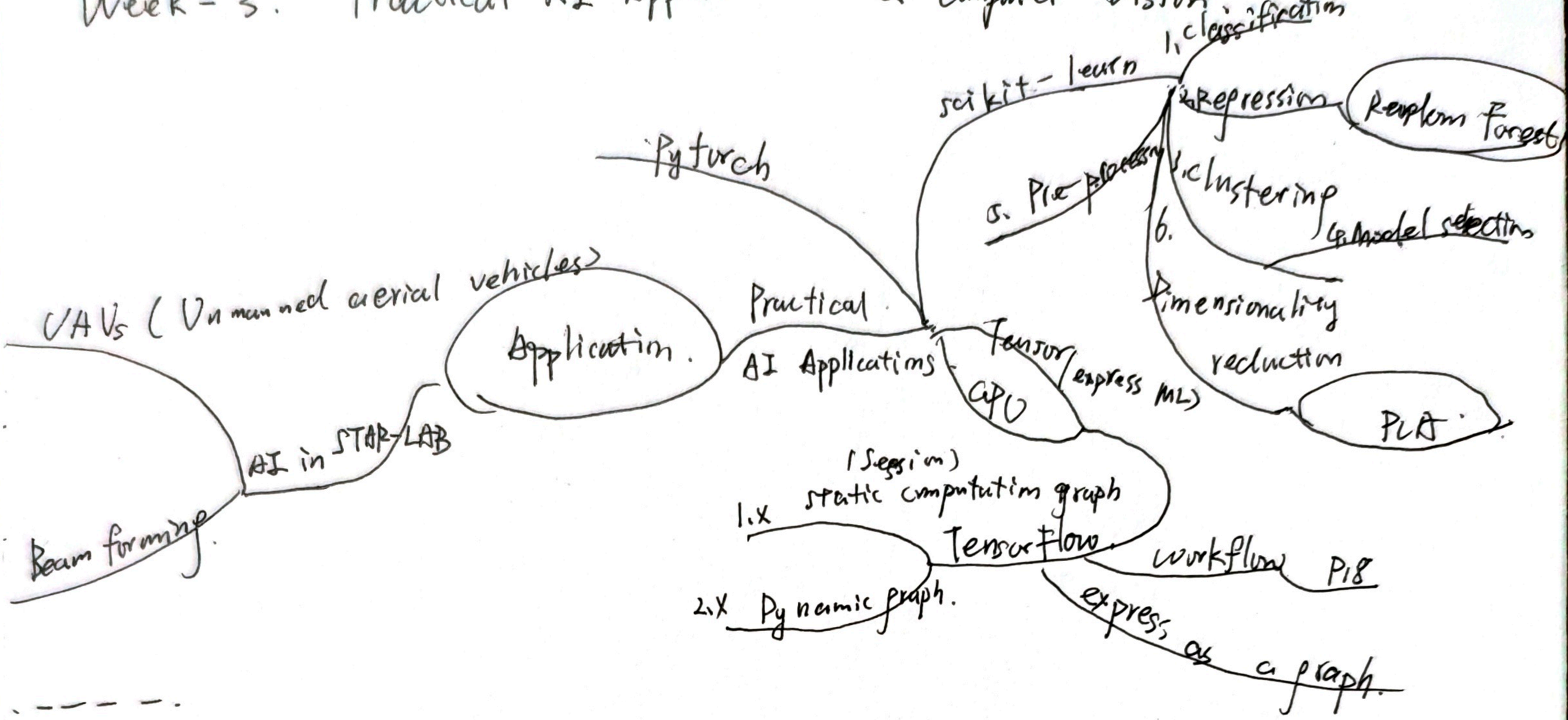
Elements of NNs.
- hidden layers
- operations
  - linear function.
  - Activation Function
    - Sigmod $\frac{1}{1+e^{-x}}$.
    - Tanh $\frac{2}{1+e^{-2x}}-1$
    - ReLU
    - Leaky ReLU
    - Softmax.

Training NNs
- Data preprocessing
  - Mean ~~subst~~ subtraction
  - Normalization
- loss function.
  - Cross-entropy.
  - Mean Squared Error $\frac{1}{N}\sum_i^N (y_i-\hat{y}_i)^2$
  - Mean Absolute Error $\frac{1}{N}\sum_i^N |y_i-\hat{y}_i|^2$.
- optimizing.
  - Gradient descent
    - Backpropagation.
    - 方式
      - Mini-batch Gradient Descent.
      - Stochastic Gradient Descent.
    - 梯度消失 (Vanish Gradient)
      - GD with Momentum.
      - Nesterov Accelerated Momentum
      - Adam.
  - Learning Rate
    - Annealing
    - Warmup
  - Vanishing gradient.
    - change learning rate.
    - ReLU, regularization.
    - LSTM.
- Generalization.
  - 过拟合 Overfitting
  - Regularization.
    - L2 weight decay $L(\theta)=L(\theta)+\lambda\Sigma|\theta|^2$.
    - L1 weight decay $L(\theta)=L(\theta)+\lambda\Sigma|\theta|$.
    - Dropout.
    - Early Stopping

NN architectures
- CNN.
- RNN
- LSTM.

- RL
  - MDP
    - Markov process.
    - MDP <S, A, R, P, γ>.
    - policy π.
    - value function → (确定每一步的action) Q-value function.
  - 分类
    - model-based
    - model-free.
    - value-based.
    - ~~value~~ policy-based
    - On-policy
    - off-policy.
  - Q-learning (使用函数拟合 $Q^*(s,a)$). 建立 Q-table
    - $Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t))$.
    - ↓
  - DQN. (因为数据量大或连续, Q-table难以建立.
  - DQN策略.
    - Policy gradient: 太多 state-action 对, 无法全部计算.
    - ↓
    - Actor-critic architecture
      - (Value-Based) + (Policy-Based).
      - ↑ ~~Value network~~ Critic network.
      - ↑ Value network.

Week-3. Practical AI Applications & Computer Vision



Q1: Explain static graph and dynamic graph used in DL frameworks.

The computation process can be view as a graph.

static graph is one that defines the entire computation graph before performing the computation. When the data is obtained, it is calculated according to the defined calculation graph.

Pymamic graphs, on the other hand, generate the computational graphs as they are computed, and the complete graph is known only when the computation is completed.

Computer Vision is the study of analysis of pictures and videos in order to achieve results similar to those as by humans.

Goal: write computer programs that can interpret images.

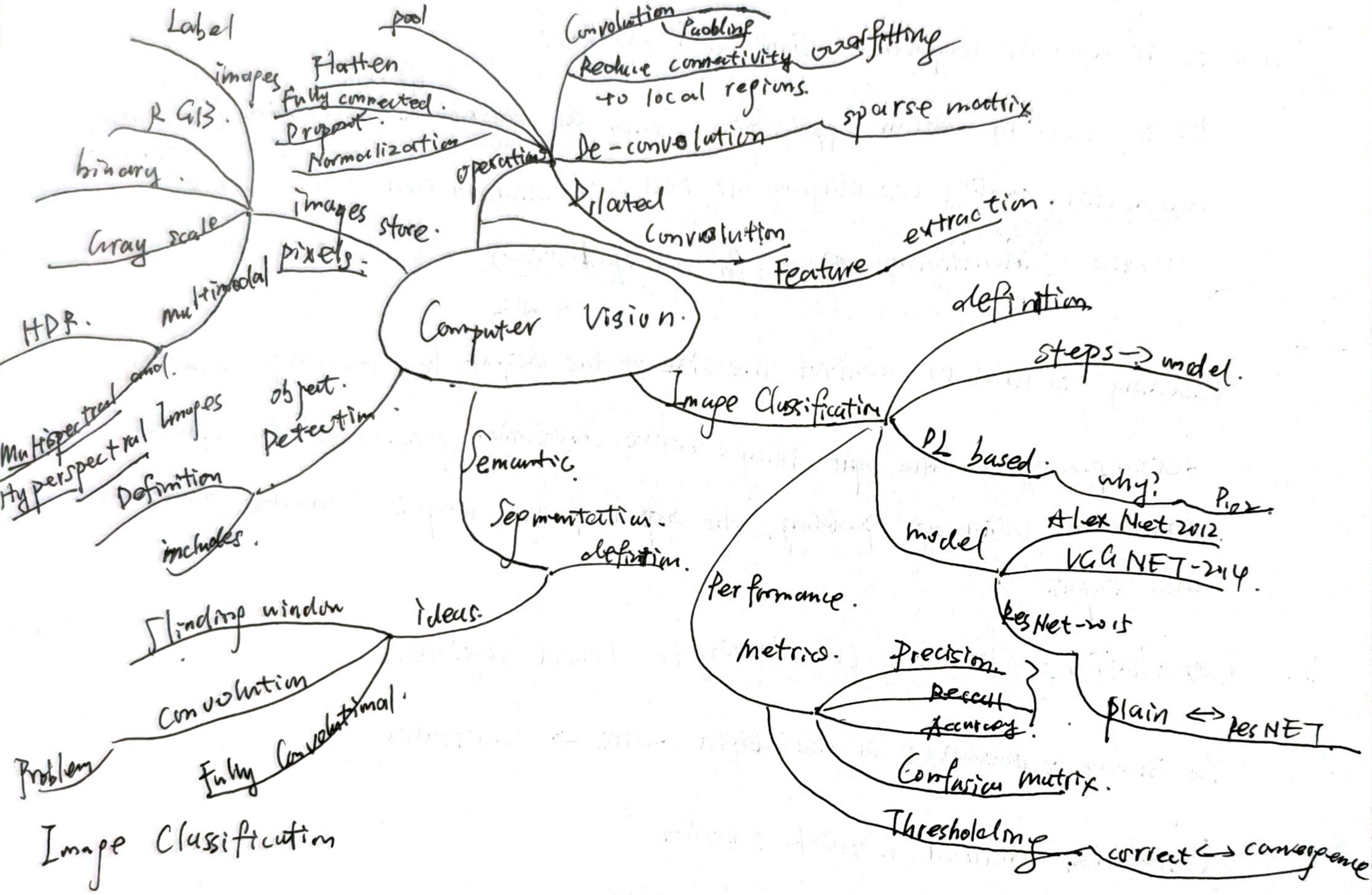


Image Classification

1. Given a set of pixels determine the category of image.

2. The fundamental steps involved in processing an image in Computer Vision.

(steps to build a computer vision model).

a. Data collection (Capturing an image).

b. Data clean (noise reduction).

c. Data preparation (Resize all the pictures to one common size)

d. Build and train the model (Code, identifying relevant characteristics. choosing important features).

...

e. make sense of visual information (classification / recognition).

Image segmentation

3. Segmention is the process of breaking an image into groups, based on similarities of the pixels.

Label each pixel in the image with a category label.

Paired training data

4. for each ~~data~~ of training data, each pixel is labeled with a semantic category.

Object Detection:

5. Definition: Object recognition is the process of identifying specific objects or instances within an image or video sequence.

6. How is it used in Computer Vision applications:

It is used in various applications, such as autonomous navigation (recognizing obstacles) augmented reality (overlaying digital information on real-world objects). robotics (identifying objects for manipulation).

7. Padding is used to control the size of the output feature map, typically zero-padding to the input image before convolution preserves its spatial dimensions. Without padding, the output feature map is smaller than the input.

8. Convolution: Reduce connectivity to local regions.

Too many parameters in the weight matrix → overfitting.

Convolution Neural network solution

- Local correlation, parameter sharing.

4. What is NLP, why is it an essential field in artifical intelligence and linguistics?

a. NLP is a vital field in AI and lingnistics that empowers computers to understand, interpret, and generate human language, facilitating more natural communication between humans and machines.

b. Its applications range from automating tasks and data understanding to multilingual communication and personalized content creation, making NLP an essential component of modern technology and information processing.

5. Unsupervised pre-training techniques.

R: The key technique of ChatGPT to achieve dialogue ~~gene~~ generation.

6. The limitations of current AI approaches.

① Data Dependency

② Interpretability and Explainability.

③ Generalization. } Domain shift.

④ Computation and Resource Requirements.

⑤ Energy Consumption.

⑥ Robustness and Security.

7. What is domain shift?

Domain shift refers to a phenomenon in machine learing and statistics where the statistical properties of data change with transitioning from one domain or distribution to another. In other word, it occurs when a model that is trained on data from one source domain performs poorly when applied to a different domain.

may include variations in viewpoint, lighting conditions, data distribution, etc.

Solution: Zero-shot learning, knowledge transfer, GAN.

8. AGI. (Artificial Generative Intelligence)

Model: from CNN-based to transformer-based.

DL+IL+RL=AGI, first imitate human data and then explore unknown ~~environ~~ environment with RL.

9. How will AGI evolve?

① LLM (Large Language model): from common to professional.

② More ~~in~~ media will be involved.

③ LLMs leverage tools to affect real world.

④ Automatic driving will be realized thanks to LLMs.

⑤ Natural language will be the new programming language.

⑥ AI for Science.

10. Example about AGI:

AlphaGo, ChatGPT, AlphaFold, ClimaX.., copilot?

一、Decision Trees.

解决分类问题的一种算法.

A simple yet effective form of learning from examples.

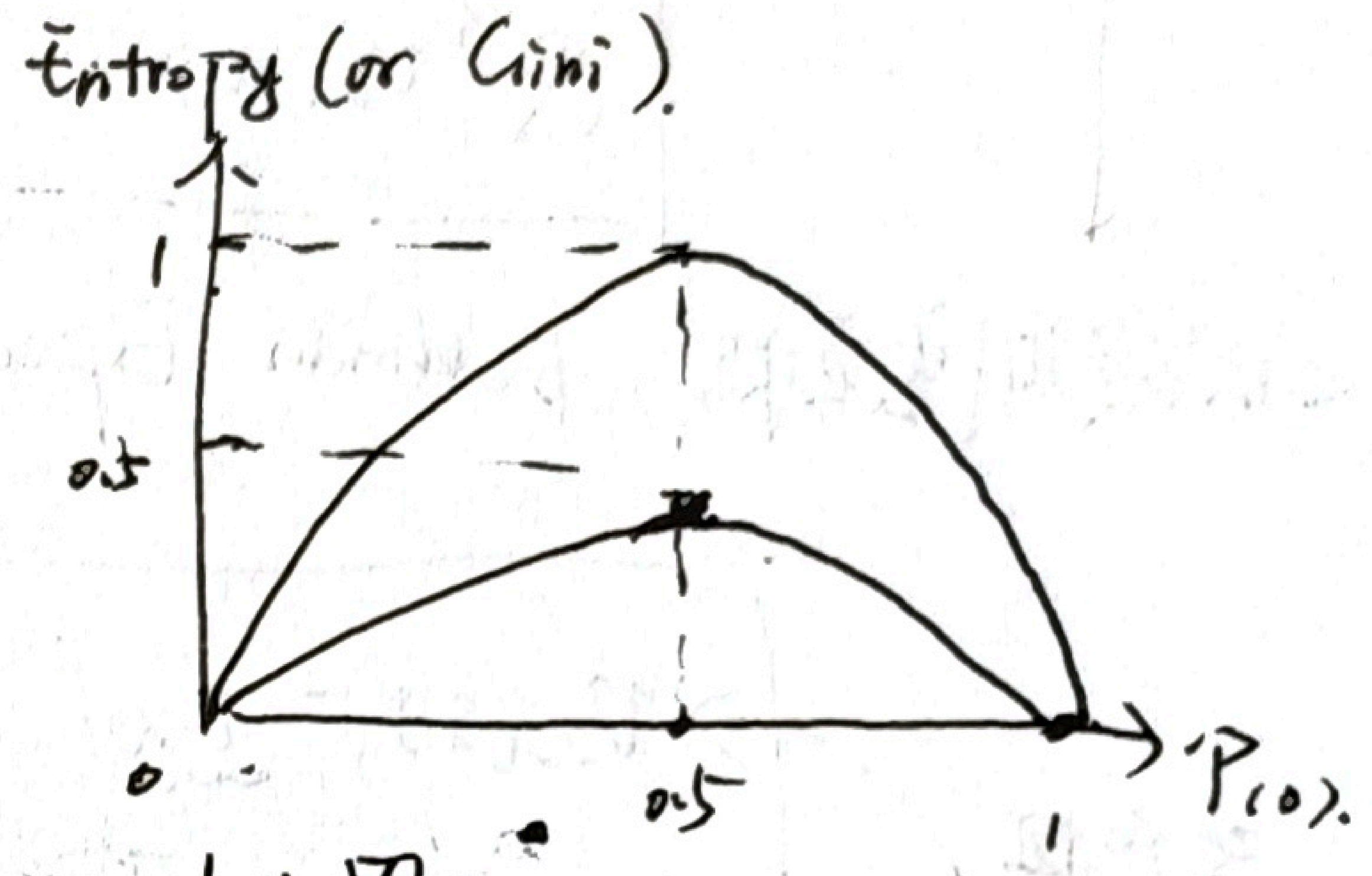
通过属性构建节点：二元，多元，连续 → 分解成二元 / 可同样选择合适属性.

选择属性

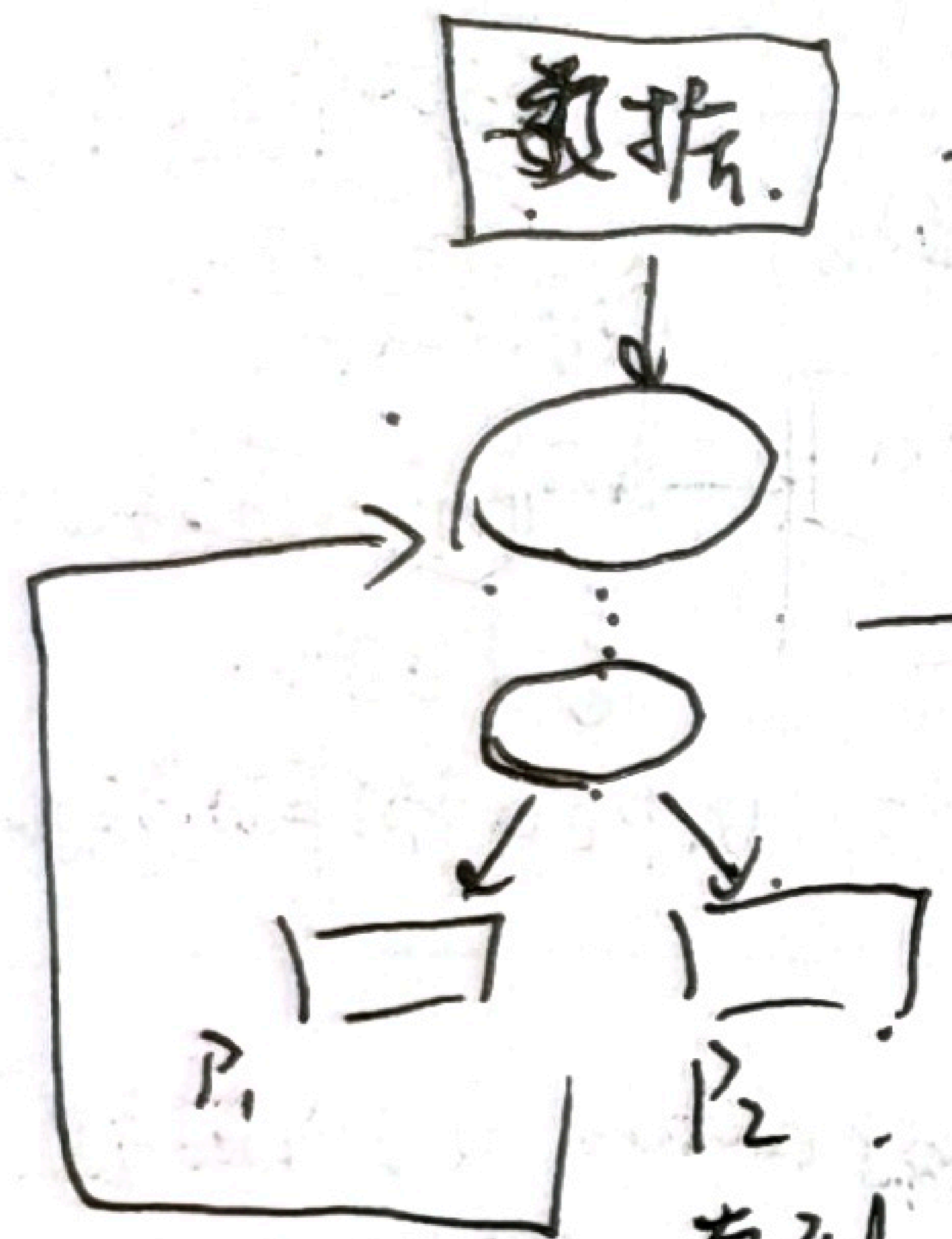
熵 — Entropy $H(x) = -\sum_{i=1}^{n} P_i \log_2 P_i$

相同等价 ⇕

(Cart算法) Gini 系数 [$Gini = 1 - \sum_{i=0}^{n} [P_i]^2$].



由此图，Entropy 或 Gini 越小，确定性越高.



数据 → $H(x) = 1$.

→ 中间计算下一节点 Gini 系数并加权比较，选择小者作为分类属性

→ $H'(x) = -\sum_{i=1}^{n} P_i \log_2 P_i$ 越小，纯度越高.

直到 output bool value.

(ID3算法) 信息增益算法. Information gain → 每到1个新的节点，H(Y)发生变化，变成当前节点对结果的信息熵.

$I(x_n, Y) = H(Y) - H(Y|x_n)$

$H(Y)$：指向最终目标. $H(Y|x_n)$：在给定条件下指向最终目标.

选择信息增益最大的属性分裂.

条件信息熵. Conditional Entropy.

$H(Y|X) = -\sum_x \sum_y P(x,y) \log_2 P(y|x) = -\sum_x P(x) \sum_y P(y|x) \log_2 P(y|x)$

每种 X_i 的发生下 Y_i 的条件信息熵之加权和

$P(x,y) = P(y|x) P(x)$.

exp:

State A	State B
A_1	B_1
A_2	B_2

↑ 一种选择. ↑ 最终判定.

下面所有的计算都指向结果

$H(B) = -P(B_1)\log_2 P(B_1) - P(B_2)\log_2 P(B_2)$.

$H(B|A_i) = -P(B_1|A_i)\log_2 P(B_1|A_i) - P(B_2|A_i)\log_2 P(B_2|A_i)$.

$H(B|A) = \sum_i P(A_i) H(B|A_i) = P(A_1)H(B|A_1) + P(A_2)H(B|A_2)$

$I(B, A) = H(B) - H(B|A)$.

所以 $I(B|A)$ 越大，说明 $H(B|A)$ 越小，在A发生的情况下（以属性A进行分类的话）纯度越高.

通过ID3算法或Cart算法逐步选择分裂节点（属性），构建决策树.