DSA 5105 LEC 10

Recap:

Image → treat one image as one dataset (each pixel is one data point)

K-means: Real-application { Compression, Segmentation } and do the clustering

→ understand the cluster of colors

→ replace the color with centroid → average

- Compression

→ replace the color with centroid

① centroid color

② assignment matrix → single channel

(raw image has 3-channels) ⇒ 1/3 compress rate

- Segmentation

→ use K-means to segment images

GMM → Density Estimation of original data set

$X \sim P(\cdot)$ $P(\cdot) \Rightarrow$ ground-truth dist.

① $\hat{P}_{Gauss} \approx P$ (estimator of Ground-truth)

→ use hidden (latent) variable

② achieved by MLE + (EM)

→ Clustering

$\Rightarrow k' = \arg\max_k P(C_k|x)$

→ Fraud Detection

Application

label-imbalanced → not a good idea to try supervised learning

Interesting: PCA can use to use GMM to do Density Estimation

transfer data safely!

↑

hidden original data!

(privacy issue)

Procedure

① data split { fraud / non-fraud

② learn the non-fraud data density through GMM

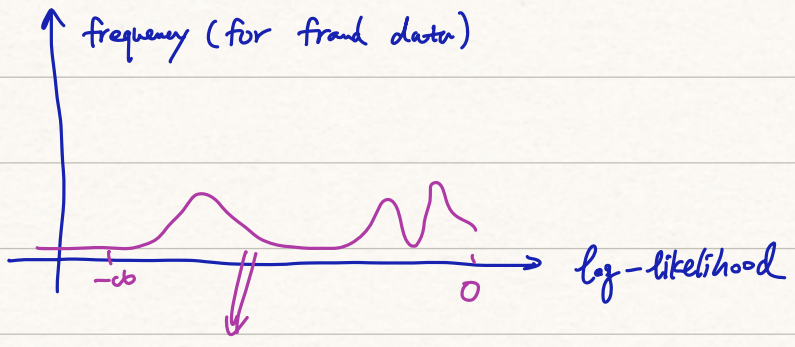
Punchline

we can use any density estimator?

③ draw the diagram

GMM fit the Non-fraud density very well

④ draw the diagram



give us the intuition that:

⇒ we can use GMM model trained by Non-fraud

calculate log-likelihood for one data

→ evidence

Performance Metric In Supervised Learning → Criterion → select threshold

{ TP / TN } { FP / FN }

Decision Rule { log-likelihood < thres, fraud / o/w, non-fraud

{ FP: Non-fraud but model says that it is fraud

FN: fraud but model says that it is non-fraud

$$\text{Precision} = \frac{TP}{TP+FP} \rightarrow \text{positive prediction}$$

$$\text{Recall} = \frac{TP}{TP+FN} \rightarrow \text{positive ground truth}$$

$$\rightarrow \text{F1-score} := 2\,\frac{\text{prex recall}}{\text{pre+ recall}}$$

Application

FN → Medical field cares more about this part

Reinforcement Learning

① Problem Setting

Agent → Action

Environment → States

→ Reward

State Space Finite

Action Space Finite

② Mathematical formulation → MDP (finite)

Action
- deterministic → perform bad
- Stochastic → perform well

0) Idea philosophy

determine on the uncertainty degree of the environment

⇓

different formulation

MC

1)
- Time step $t=0,1,2,\ldots$
- State Space S → Finite MC, Assume $|S| < \infty$
- State: $S_t \in S$

$\Rightarrow \{S_t\}_{t\geq 0}$ → stochastic process

$\Rightarrow$ Transition Probability: $P(S_{t+1} = s' \mid S_t = s, \ldots)$

Markov Property (simplification)

$$\mathbb{P}(S_{t+1}=s' \mid S_t=s, \dots) = \mathbb{P}(S_{t+1}=s' \mid S_t=s)$$

$$= P_t(s' \mid s)$$

Time homogeneity (simplification)

$$\mathbb{P}(S_{t+1}=s' \mid S_t=s, \dots) = P_{ss'} := \mathbb{P}(S_{t+1}=s' \mid S_t=s)$$

$P_{ss'}$ → Transition Matrix $P \in \mathbb{R}^{n \times n}$

2) MDP

- Sequence of Time: $t = 0, 1, 2, \dots$
- State Space: $S_t \in \mathcal{S}$
- Action Space: $A(S_t) \overset{A_t}{\in} \mathcal{A}(S_t) \subseteq \mathcal{A}(\mathcal{S})$
- Reward: $R_{t+1} \in \mathbb{R}$
  - depends on S_t & A_t

MDP Trans. Prob. $:= \mathbb{P}(S_{t+1}=s', R_{t+1}=r \mid S_t=s, A_t=a)$

$:= p(s', r \mid s, a)$

if we have the policy then R_{t+1} only depends on state S_t

A MDP is finite $\Leftrightarrow$ $\begin{cases} |\mathcal{S}| < \infty \\ |\mathcal{A}| < \infty \end{cases}$

Decision $\longleftrightarrow$ Policy

→ the only degree of freedom for Agent to control the system

{ ① Stochastic : $\pi(a|s) = \mathbb{P}(A_t = a \mid S_t = s)$

② deterministic : $\pi : \mathcal{A} \longrightarrow S$ → not applied in real life

Returns : $G_t = \sum_{i=0}^{+\infty} \gamma^i R_{t+i+1}$ $\gamma \rightarrow$ discount rate

3) Goal of RL

$$\pi^* = \arg\max_{\pi} \mathbb{E}^{\pi}[G_0 \mid S_0 = s]$$

(G_0: return from to; start from some state $s \in S$)

$$= \arg\max_{\pi} \mathbb{E}^{\pi}[R_1 + \gamma R_2 + \dots \mid S_0 = s]$$

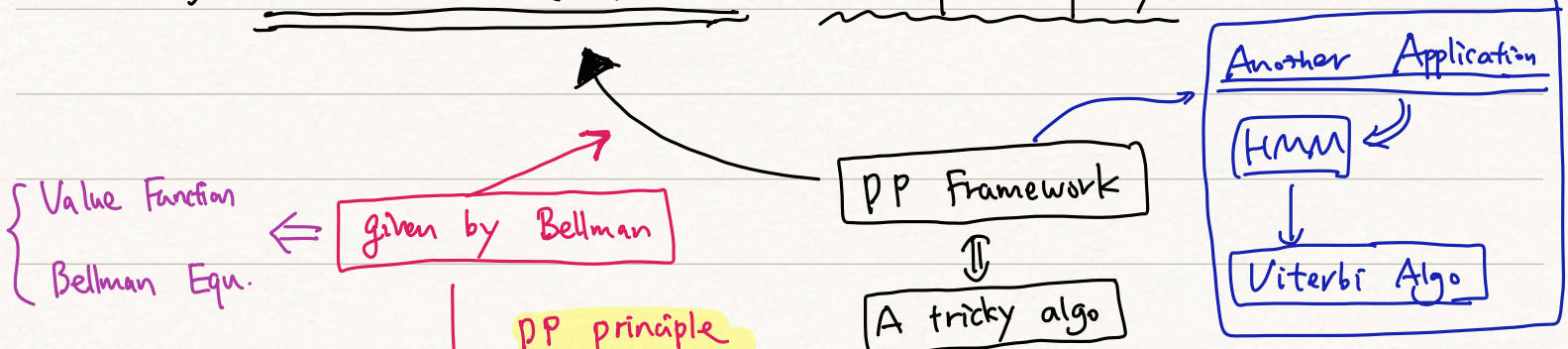
(R_1, R_2: reward, reward)

Observation :

→ if we use a Time-homogeneous MC,

then $\mathbb{E}^{\pi}[G_0 \mid S_0 = s] = \mathbb{E}^{\pi}[G_t \mid S_t = s]$

4) How to Solve this ? (find optimal policy)



$O(N^2T)$

- global optimal $\Rightarrow$ sub-optimal (most important idea)

  → decompose the question backwards