Prediction vs. Decision Making

Decisions ≡ Actions ≡ Choices   eg. Steer left/right/not (autonomous driving)

Algorithmic approach based on data

recommendation systems (movies, products, reels,

games (bots)

triaging decisions - healthcare
\ 911

What we'll cover

- Methods + Theory behind autonomous decision making

  Experimental design
  Active learning
  Optimization (statistical)
  Bandits
  RL

- focus on theoretical foundations

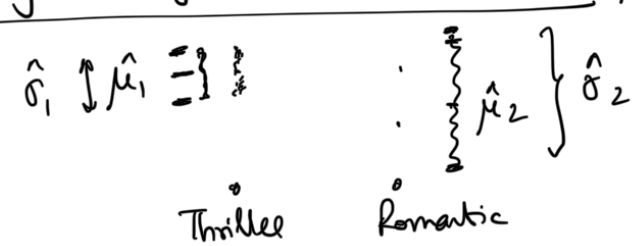
What we'll NOT cover

- Implementations (Deep RL 10-703)
- Ad-hoc
- Human-AI complementarity in Decision Making (xx)
- Ethical considerations (Hoda, 10-735 Responsible AI
  ML ethics + society 713, 7[illegible]
  Fairness 712)
- Adversarial

Naive approach

Predict using supervised learning
Recommend choice with highest reward
] Greedy

Uncertainty or Confidence in Predictions



$E[(\hat{f}(x) - E[\hat{f}(x)])^2]$ regression

• variance $E[(f(x) - E[f(x)])^2]$
predictor

• $P(Y=1|X)$ - logistic, NN? classification

Performance Measure for Decision Making

Input $x \equiv \{s, a\}$ → a - actions, s - state

Output $y \equiv r$ reward $E[r] = \mu$ mean reward

① Best decision / choice $x^* \equiv a^*$ best action $= \arg\max_a \mu(a)$

$\pi^*(a|s)$ best action for each state

$$\min_{x_1..x_T} E[|x^* - \hat{x}_T|]$$

↳ x_T or $\sum_{t=1}^{T} x_t$ or ...

② Simple regret

$$\min_{x_1..x_T} E[\mu(x^*) - \mu(\hat{x}_T)] = \mu(x^*) - E[\mu(\hat{x}_T)]$$

① + ② exploration only — { randomly, adaptively

③ Cumulative regret

$$\rightarrow \min_{x_1..x_T} \sum_{t=1}^{T} E[\mu(x^*) - \mu(x_t)]$$

exploration + exploitation

low confidence actions — most promising actions

Exp design 1-shot decision under data budget k out of n choices of data points.

$$\min_{x_1..x_k} Risk(\hat{f}_{x_1..x_k}) - Risk(f^*) \equiv ②$$

Active learning sequentially chose k data points out of n

$\equiv$ ②

Optimization $\quad x \to r(x)$

1 or 2

sequentially choosing action that max reward

Bandits $\quad$ sequentially chose actions that max cumulative reward 3

$x \equiv a$

Contextual Bandits $\quad$ — do — 3

$x \equiv (a, s)$

↳ context (independent)

Reinforcement Learning $\quad$ — do — 3

$x \equiv (a, s)$

↳ state (dependent on past)

$$s_{t+1} \leftarrow \{s_\tau, a_\tau\}_{\tau=1}^{t}$$

Concentration Bounds

wp $\geq 1-\delta \quad |\theta^* - \hat{\theta}| \leq \varepsilon$.

$\theta^* \in (\hat{\theta} \pm \varepsilon)$

$\hat{\theta} \overset{iid}{\sim} x_1 \ldots x_T$

$\hat{\theta} \leftarrow$ dependent data

$x_1 \to x_2 \to x_3 \ldots$

$R(j^*) - R(\hat{j})$