Welcome

Kevin Jamieson (Instructor)

Lalit Jain (co-instructor)

Yuhao Wan (TA)

Interactive learning in ~~non-stochastic~~ better modeled environments

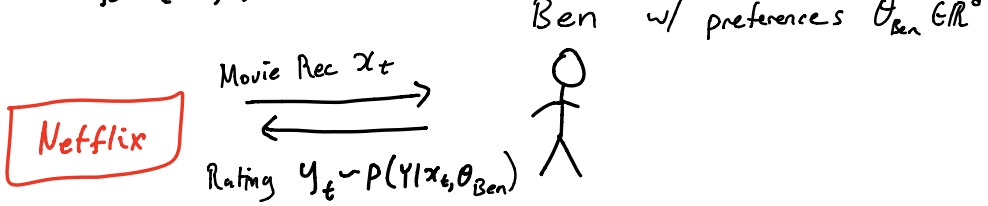
Machine learning relies on stochastic IID assumptions

$\{(x_i, y_i)\}_{i=1}^n$ where $(x_i, y_i) \overset{iid}{\sim} \theta_{xy}$

Learns predictor $\hat{h} : \mathcal{X} \to \mathcal{Y}$ using $\{(x_i, y_i)\}_{i=1}^n$

so that for new point $(x_a, y_a) \overset{iid}{\sim} \theta_{xy}$ $\hat{h}(x_a) \approx y_a$.

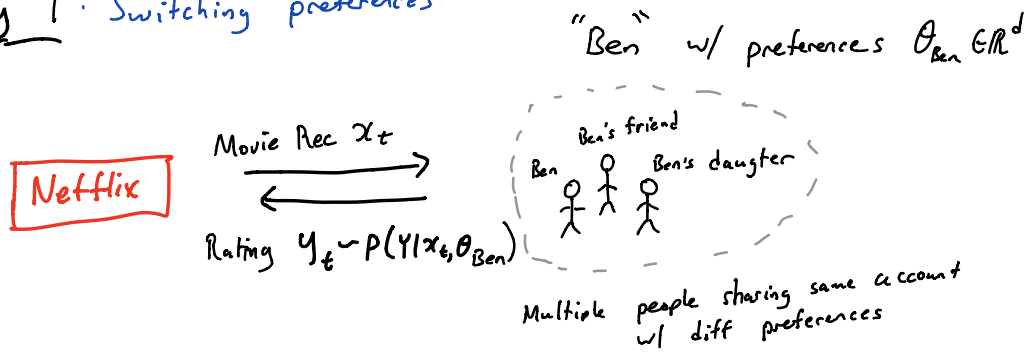
Model for $t = 1, 2, \ldots$



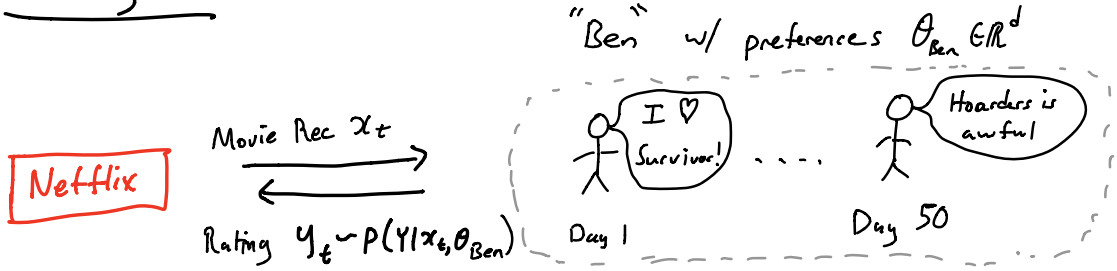
Things we worry about

- Trading off exploration versus exploitation (i.e. don't be too greedy)

Reality 1: Switching preferences



Reality 2: Time variation



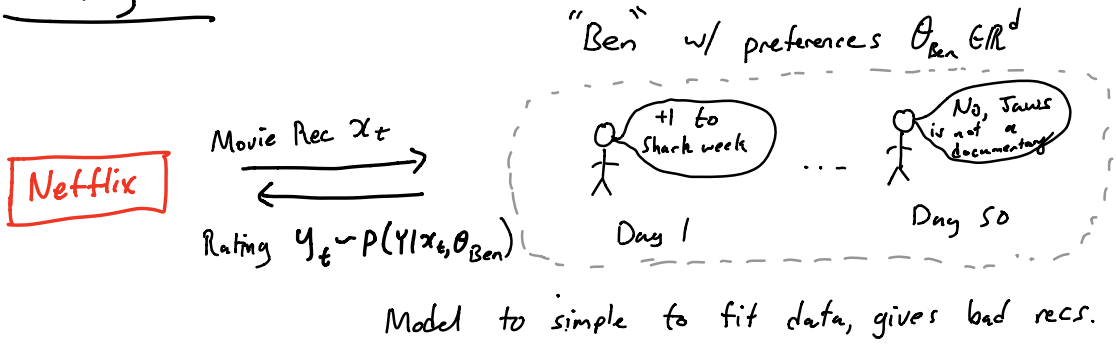
User preferences change over time...

Content producers react to changing preferences...

User preferences react to changing content...

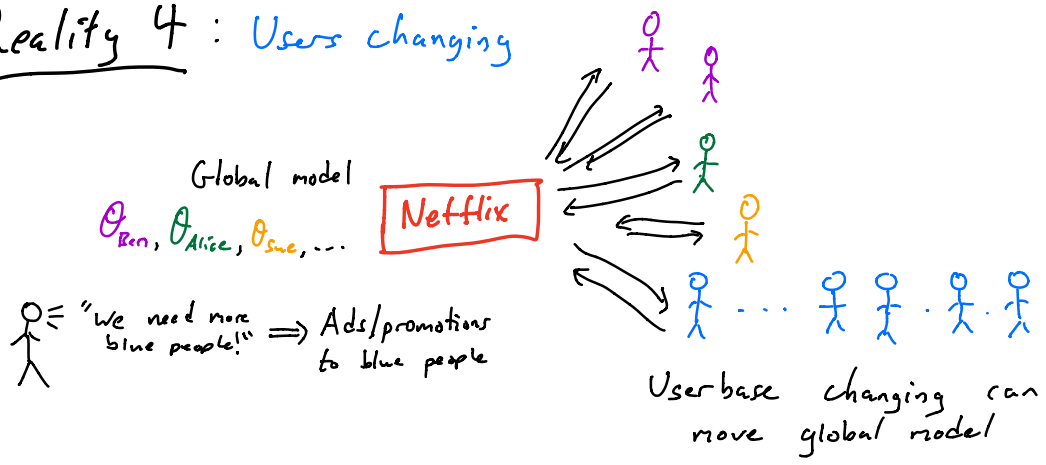
⋮

Reality 3 : Model misspecification



Model to simple to fit data, gives bad recs.

Reality 4 : Users changing



This class is about

- Modeling
- Robustness
- Guarantees
- Practical algorithms (mostly)

Inspired by real-world scenarios we have encountered.

Class Logistics, evaluation, plan

zoom etiquette

Multi-armed Bandit

Each "arm" is a source s.t. when "pulled" emits some observation $X \in \mathbb{R}$

$n \geq 2$, $[n] = \{1, 2, \ldots, n\}$

Input n arms

for $t = 1, 2, \ldots, T$

$X_t \in \mathbb{R}^n$

Nature chooses $X_{t,i} \in \mathbb{R}$ for $i = 1, 2, \ldots, n$ (secretly)

Player chooses arm $I_t \in [n]$ and observes X_{t, I_t}

The regret of the player is defined as

$$R(T) = \max_{i \in [n]} \sum_{t=1}^{T} X_{t,i} - \sum_{t=1}^{T} X_{t, I_t}$$

The reward of playing the single best action in hindsight

reward obtained by the player

Typically, goal is to define an algorithm s.t. $R(T) = o(T)$ so that $\frac{R(T)}{T} \to 0$

If we assume nothing about rewards at all

then $\exists$ choice of rewards $\{X_t\}_t \subset \mathbb{R}^n$ s.t.

$R(T) = \Omega(T)$ w. const prob.

So, we make assumptions.

Adversarial setting:

Bounded rewards $X_{t,i} \in [0,1] \quad \forall t,\ i \in [n]$

$\exists$ algorithm s.t. $\mathbb{E}[R(T)] \leq O(\sqrt{nT})$

Stochastic setting

$\exists$ distributions over $\mathbb{R}$ s.t. $\nu_1, \ldots \nu_n$

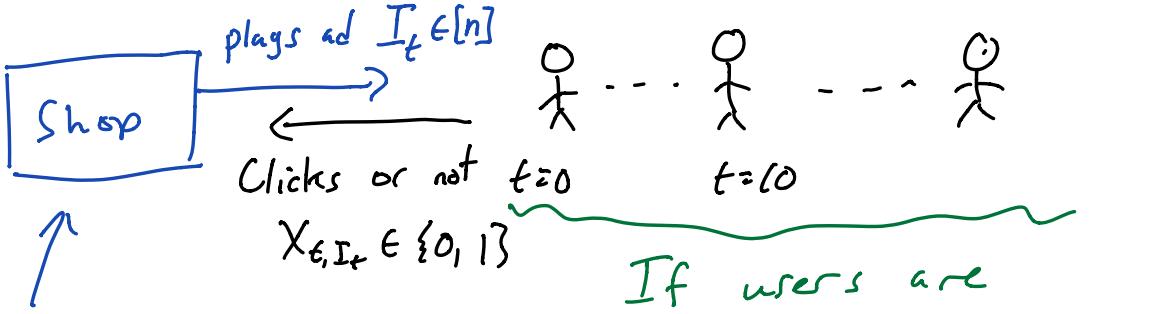
$X_{t,i} \overset{iid}{\sim} \nu_i$ for all $t \in \mathbb{N}$.

Typically we assume each ν_i has bounded support, 1st/2nd moment, or is sub-Gaussian.

$\exists$ algorithm s.t. $\mathbb{E}[R(T)] \leq O(\min\{C_\nu \log(T), \sqrt{nT}\})$

where C_ν depends on "hardness" of $\nu_1, \ldots, \nu_n$

When is stochastic justified? (Not on T)



Shop is player choosing which arm

If users are independent then at each time t $x_{t,i} \sim \nu_i$