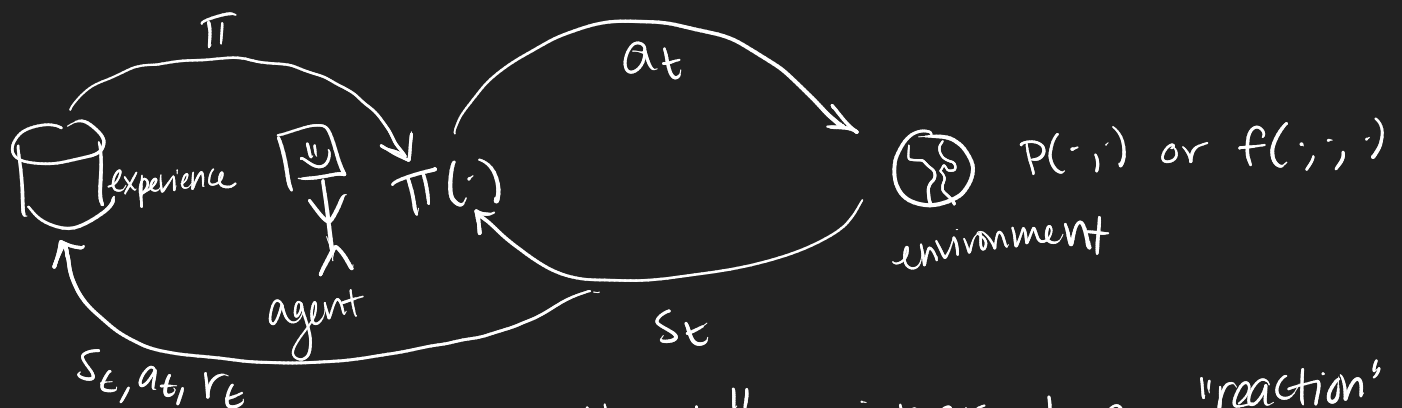
1) Types of Feedback in RL

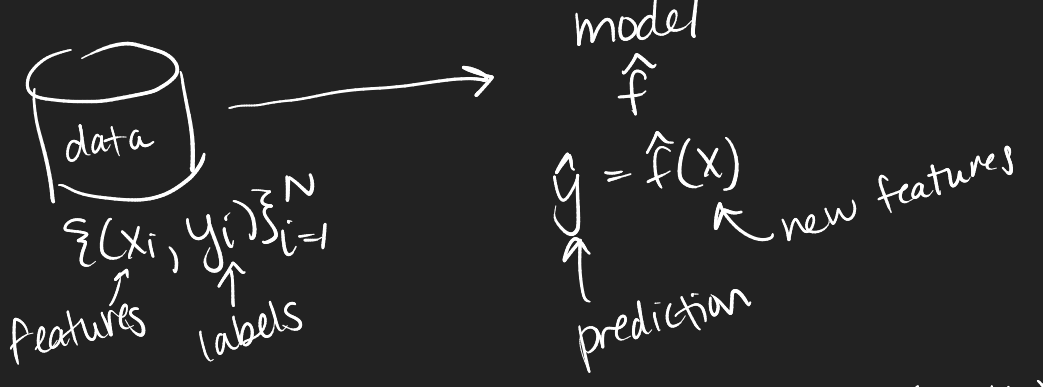


1) "control feedback" – inner loop "reaction"
   ex- thermostat regulating temperature
   Unit 1
2) "data feedback" – "outer" loop "adaptation"
   ex- smart thermostat learn preferences
   Unit 27

From now on, transition / dynamics are <u>unknown</u>
$P(\cdot,\cdot)$ $f(\cdot,\cdot,\cdot)$

(often $r(\cdot,\cdot)$ also unknown)
reward

2) Supervised Learning (SL)



e.g. classification of images
x - image (pixels)
y - cat vs. dog

regression
x- financial history
y- probability of loan repayment

Analogy between SL and special case RL

$$\min_f \mathbb{E}_{(x,y)\sim D}\left[\ell(y,\hat{y}) \mid \hat{y} = f(x)\right]$$

features/states
predictions/actions
model / policy
offline / online
no impact
dependence
loss/cost
distribution / transition probabilities
iid
dependent

What should we learn to solve (i.e. find near-optimal policy)

$$\mathcal{M} = \{S, \mathcal{A}, \underbrace{P, r}_{\text{unknown}}, \gamma\}$$

- transition P, reward r
- value function V^π, Q function Q^π } plug in π & observe
- optimal value/Q function V^*, Q^*
- optimal policy π^*

supervision

✓ $s_{t+1} \sim P(s_t, a_t)$, $r_t = r(s_t, a_t)$

✓≈ after delay $\sum_0^t \gamma^t r_t$

✗

✗ exception: imitation learning

3) Estimation & Prediction

A) Tabular Setting:

suppose $x_i \overset{iid}{\sim} D$, $x_i \in \mathcal{X}$, $p(x) = P_D(x_i = x)$

$$\hat{p}(x) = \frac{1}{N}\sum_{i=1}^N \mathbb{1}\{x_i = x\}$$

How good is $\hat{p}$?

Lemma: (Consistency)

$$\mathbb{E}_{x_1,\dots,x_N}[\hat{p}(x)] = p(x)$$

Proof: $\mathbb{E}[\hat{p}(x)] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}[\mathbb{1}\{x_i = x\}]$

$= \frac{1}{N}\sum_{i=1}^{N} P(x_i = x)$

$= p(x)$

Theorem (concentration)

For all $x \in \mathcal{X}$, with probability $1-\delta$

"pointwise" $|\hat{p}(x) - p(x)| \leq \sqrt{\frac{2\log(2|\mathcal{X}|/\delta)}{N}} \approx O\left(\frac{1}{\sqrt{N}}\right)$

Proof out of scope (Hoeffding's Inequality)

$x, y \sim D$ $\quad \hat{f}(x) \approx y$ $\quad \{(x_i, y_i)\}_{i=1}^{N}$

By similar approach,

$$\hat{f}(x) = \frac{\sum_{i=1}^{N} y_i \mathbb{1}\{x = x_i\}}{\sum_{i=1}^{N} \mathbb{1}\{x = x_i\}}$$

Details out of scope, but if $y = f^*(x) + w$ (w: iid noise)

$\forall x \in \mathcal{X}$, w.p. $1-\delta$

pointwise $|\hat{f}(x) - f^*(x)| \lesssim \sqrt{\frac{|\mathcal{X}|\log(1/\delta)}{N}}$

applies when $|\mathcal{X}| < N$

B) non-tabular:

Empirical Risk Minimization

$$\hat{f} = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N} \ell(y_i, f(x_i))$$

$\mathcal{F}$: function class; ℓ: loss; y_i: label; $f(x_i)$: prediction

1) Estimation (Parametric)

$\mathcal{F} = \{ f_\theta(x) \mid \theta \in \mathbb{R}^d \}$

eg. neural network θ - weights

e.g. $f_\theta(x) = \theta^T \phi(x)$ ($\phi(x)$: fixed)

Suppose $y = f_{\theta^*}(x) + w$ (w: iid noise)

(θ^*: true parameter)

$\hat{f} = f_{\hat{\theta}}$

Estimation Error: $\|\theta^* - \hat{\theta}\|$

Details out of scope, estimation error bounded

parametric $\|\theta^* - \hat{\theta}\| \lesssim \sqrt{\frac{d \log(1/\delta)}{N}}$

($\lesssim$: hiding constants)

useful when $N > d$

Prediction

expected prediction error on $(x, y) \sim D$

$$\mathbb{E}_{x,y \sim D}[\ell(\hat{f}(x), y)]$$

Assume D: $x \sim D_x$, $y = f^*(x) + w$ (w: noise) and $f^* \in \mathcal{F}$

Details out of scope, but

$$\mathbb{E}_{x,y \sim D}[\ell(\hat{f}(x), y)] \lesssim \sqrt{\frac{\log(1/\delta)}{N}}$$

Prediction error is about average case and fixed distribution