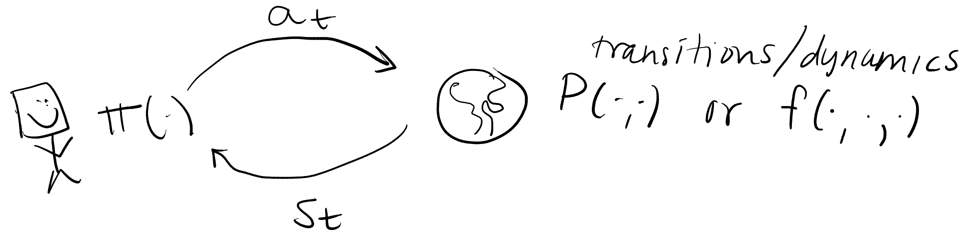
CS 4/5789 21 Feb 2022 Prof Sarah Dean

Lecture 9: Prediction and Estimation

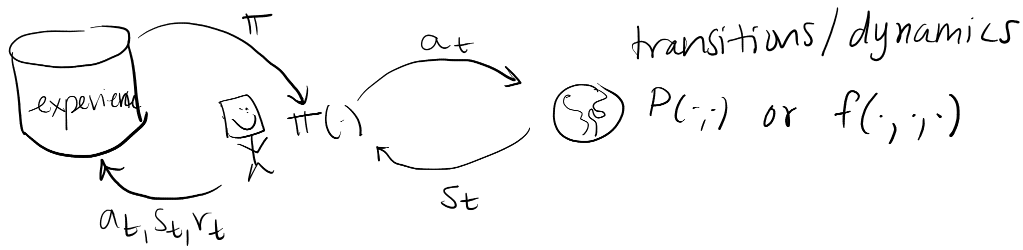
1) Types of Feedback in RL

1) control feedback "reaction"



- feedback between states & actions
- historically studied in control theory
  "automatic feedback control"
  ex - thermostat regulates temperature
- we focused on this level for unit 1

2) Data Feedback "adaptation"

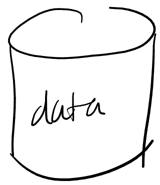


- feedback between policy and data
- connections to machine learning
  ex - smart thermostat learns preferences
- we consider this level starting in Unit 2

From now on: the transitions/dynamics $P(\cdot,\cdot)$ or $f(\cdot,\cdot,\cdot)$ are <u>unknown</u>. (often also the reward $r(\cdot,\cdot)$)

2) Supervised learning

Recall the process of learning predictive models from data



$\longrightarrow$ model f

$\{(x_i, y_i)\}_{i=1}^{N}$ — x_i: features, y_i: labels

$\hat{y} = f(x)$ — $\hat{y}$: predicted label, x: new features

e.g. classification: x: image
y: cat or dog

We can actually view supervised learning as special case of reinforcement learning where control feedback doesn't matter because "actions" (predictions) do not impact the environment.

SL	$\longleftrightarrow$	RL special case
features x		states s
predictions $\hat{y}$		actions a
model f		policy π
data distribution $\mathcal{D}$		transition probability $P(s,a)$
loss (accuracy) $\ell(y, \hat{y})$		cost (reward) c

$$\min_{f} \; \mathbb{E}_{(x_i, y_i) \sim \mathcal{D}} \left[\ell(y_i, \hat{y}_i) \;\middle|\; \hat{y}_i = f(x_i) \right]$$

Traditional supervised learning does not typically consider a time horizon or the problem of exploration. We will explore this aspect further by studying "bandit problems" in Unit 5. Nevertheless, supervised learning is the foundation of data feedback in RL.

What might we use SL to learn?

- the "model": transitions $P(\cdot,\cdot)$ or dynamics $f(\cdot,\cdot,\cdot)$ (& rewards $r(\cdot,\cdot)$)
- value of some policy π $V^\pi(\cdot)$ and $Q^\pi(\cdot,\cdot)$
- optimal value: $V^*(\cdot)$ and $Q^*(\cdot,\cdot)$
- optimal policy $\pi^*(\cdot)$

Are we able to supervise the above learning problems? (e.g. observe the labels)

- model: yes, at the next timestep
- value of π: sort of, at the end of the time horizon (or approx. with discounting)
- optimal value: not directly
- optimal policy: not directly, unless we have expert demonstrations (imitation learning)

preview of the challenges to come.

3) Estimation And Prediction

Since Supervised Learning is an important foundation for RL, we will recap / discuss some important results.

A) Tabular Setting: counting

Let $X \in \mathcal{X}$ be distributed according to $\mathcal{D}$, and let $p(x) = \mathbb{P}(X = x \mid X \sim \mathcal{D})$.

Suppose $\mathcal{D}$ is unknown but we have a set of samples $\{X_i\}_{i=1}^{N}$.

Empirical (estimated) distribution:

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

How good is this estimate?

Lemma (consistency):

$$\mathbb{E}_{X_1, \ldots, X_N}(\hat{p}(x)) = p(x)$$

expectation over random sample

Proof:

$$\mathbb{E}_{X_1, \ldots, X_N}(\hat{p}(x)) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{X_i}\left[\mathbb{1}\{X_i = x\}\right]$$ (linearity of expectation)

$$= \mathbb{E}_{X_i}\left[\mathbb{1}\{X_i = x\}\right]$$ (X_i are identically distributed)

$$= \mathbb{P}\{X_i = x\}$$ (The expectation of indicator on event is equal to the probability of the event)

$$= p(x)$$ (definition)

Lemma (concentration)

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

$$|\hat{p}(x) - p(x)| \leq \sqrt{\frac{2\log\left(\frac{2|\mathcal{X}|}{\delta}\right)}{N}}$$

Proof: out of scope, but uses "Hoeffding's inequality"

By similar logic, we can generalize from probability estimation to prediction by

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

average of values observed in data

Details out of scope, but if $y = f^*(x) + w$ (iid bounded noise) we can often derive a bound like

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

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

But this doesn't work well when the size of $\mathcal{X}$ gets large compared to # samples

B) Non-tabular setting

Suppose $x, y \sim \mathcal{D}$, data $\{(x_i, y_i)\}_{i=1}^N$ and we want to learn a map $\hat{f}$ which predicts y from x.

Empirical Risk Minimization

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

class of functions we consider

loss function

prediction

label

1) Parameter Estimation

often, class of functions $\mathcal{F}$ is parametric:

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

e.g. neural network w/ fixed architecture, θ represents weights

e.g. $f_\theta(x) = \theta^T \phi(x)$ ← known transformation

Supposing that the labels y are generated by some true parameter θ_*

$$y = f_{\theta_*}(x) + w$$

iid noise

We can evaluate learned model $f_{\hat{\theta}}$ by closeness to true parameter:

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

Details are out of scope, but often, the estimation error can be bounded by (with probability $1-\delta$)

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

need # samples to be much larger than parameter dimension

Example: least-squares

Let $y = \theta_*^T \phi(x) + w$ with $w \sim \mathcal{D}$ i.i.d. noise, $\theta_* \in \mathbb{R}^d$ some unknown parameter, and $\phi: \mathcal{X} \to \mathbb{R}^d$ some known featurization.

Suppose dataset $\{(x_i, y_i)\}_{i=1}^N$. Then least squares estimation:

$$\hat{\theta} = \arg\min_\theta \sum_{i=1}^N (\theta^T \phi(x_i) + y_i)^2$$

We can write out the form of $\hat{\theta}$ in terms of data matrices:

$$\underbrace{\Phi = \begin{bmatrix} \phi(x_1)^T \\ \vdots \\ \phi(x_N)^T \end{bmatrix}}_{N \times d} \qquad \underbrace{Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{N}$$

$$\hat{\theta} = \arg\min_\theta \|\Phi\theta - Y\|_2^2 = (\Phi^T\Phi)^{-1}\Phi^T Y$$

2) Prediction

Another way to evaluate $\hat{f}$ is its expected prediction error on a new sample $(x,y) \sim \mathcal{D}$

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

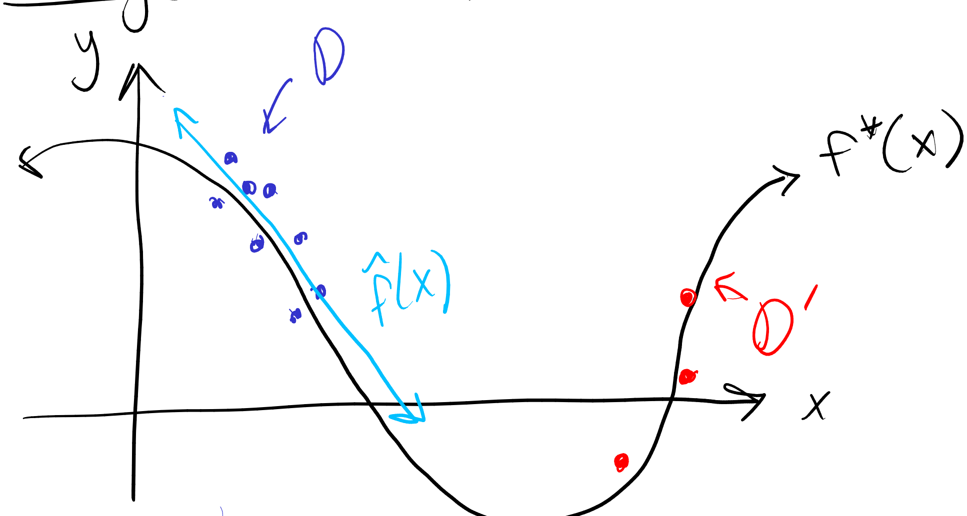
measure error by loss

Often we assume that $x \sim \mathcal{D}_X$ and $y = f^*(x) + w$ where $f^* \in \mathcal{F}$ (called realizability)

Again, the details are out of scope, but often the prediction error can be bounded w.p. $1-\delta$

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

However, prediction error guarantees only average case performance on distribution $\mathcal{D}$.



A model $\hat{f}$ learned on $\mathcal{D}$ may perform badly on some new $\mathcal{D}'$