Bayesian Networks
+
Reinforcement Learning:
Markov Decision Processes
Reminders

• Homework 7: Hidden Markov Models
  – Out: Fri, Mar. 31
  – Due: Mon, Apr. 10 at 11:59pm
GRAPHICAL MODELS: DETERMINING CONDITIONAL INDEPENDENCIES
What Independencies does a Bayes Net Model?

- In order for a Bayesian network to model a probability distribution, the following must be true:
  
  Each variable is conditionally independent of all its non-descendants in the graph given the value of all its parents.

- This follows from \( P(X_1, \ldots, X_T) = \prod_{t=1}^{T} P(X_t \mid \text{parents}(X_t)) \)

- But what else does it imply?

Slide from William Cohen
What Independencies does a Bayes Net Model?

Three cases of interest...

Cascade

Common Parent

V-Structure
What Independencies does a Bayes Net Model?

Three cases of interest...

Cascade

\[ X \perp Z \mid Y \]

Knowing Y decouples X and Z

Common Parent

Knowing Y couples X and Z

V-Structure

Knowing Y couples X and Z
Whiteboard

Proof of conditional independence

X ⊥ Z | Y

(The other two cases can be shown just as easily.)
The “Burglar Alarm” example

- Your house has a twitchy burglar alarm that is also sometimes triggered by earthquakes.
- Earth arguably doesn’t care whether your house is currently being burgled.
- While you are on vacation, one of your neighbors calls and tells you your home’s burglar alarm is ringing. Uh oh!

Quiz: True or False?

\[ \text{Q1} \quad \text{Burglar} \parallel \text{Earthquake} | \text{PhoneCall} \]

- 30% \( A = \text{t} \)oxic
- 70% \( B = \text{True} \)
- 70% \( C = \text{False} \)

“explaining away”
The “Burglar Alarm” example

• After you get this phone call, suppose you learn that there was a medium-sized earthquake in your neighborhood. Oh, whew! Probably not a burglar after all.
• Earthquake “explains away” the hypothetical burglar.
• But then it must **not** be the case that

\[ \text{Burglar} \perp\!\!\!\!\!\!\!\perp \text{Earthquake} \mid \text{Phone Call} \]

even though

\[ \text{Burglar} \perp\!\!\!\!\!\!\!\perp \text{Earthquake} \]
Def: the co-parents of a node are the parents of its children.

Def: the Markov boundary of a node is the set containing the node’s parents, children, and co-parents.
**Markov Boundary**

**Def:** the **co-parents** of a node are the parents of its children.

**Def:** the **Markov boundary** of a node is the set containing the node’s parents, children, and co-parents.

**Example:** The Markov boundary of $X_6$ is \{$X_3, X_4, X_5, X_8, X_9, X_{10}$\}
Markov Boundary

**Def:** the co-parents of a node are the parents of its children.

**Def:** the Markov boundary of a node is the set containing the node’s parents, children, and co-parents.

**Theorem:** a node is conditionally independent of every other node in the graph given its Markov boundary.

**Example:** The Markov boundary of $X_6$ is $\{X_3, X_4, X_5, X_8, X_9, X_{10}\}$.
D-Separation

**Definition #1:**
Variables X and Z are **d-separated** given a **set** of evidence variables E (variables that are observed) iff every path from X to Z is “blocked”.

A path is “blocked” whenever:

1. \( \exists Y \) on path s.t. \( Y \in E \) and Y is a “common parent”
   
   ![Diagram](image)

2. \( \exists Y \) on path s.t. \( Y \in E \) and Y is in a “cascade”
   
   ![Diagram](image)

3. \( \exists Y \) on path s.t. \{Y, descendants(Y)\} \( \notin E \) and Y is in a “v-structure”
   
   ![Diagram](image)

**If** variables X and Z are **d-separated** given a **set** of variables E  
**Then** X and Z are **conditionally independent** given the **set** E
If variables $X$ and $Z$ are \textbf{d-separated} given a \textbf{set} of variables $E$ 
Then $X$ and $Z$ are \textbf{conditionally independent} given the \textbf{set} $E$

\textbf{Definition #2:}\nVariables $X$ and $Z$ are \textbf{d-separated} given a \textbf{set} of evidence variables $E$ \textbf{iff} there does \textbf{not} exist a path between $X$ and $Z$ in the \textbf{undirected ancestral moral graph} \textbf{with E removed}.

1. **Ancestral graph**: keep only $X$, $Z$, $E$ and their ancestors
2. **Moral graph**: add undirected edge between all pairs of each node’s parents
3. **Undirected graph**: convert all directed edges to undirected
4. **Givens Removed**: delete any nodes in $E$

\textbf{Example Query:} \quad $A \perp B \mid \{D, E\}$

\begin{itemize}
  \item Original:
  \begin{itemize}
    \item A
    \item B
    \item C
    \item D
    \item E
    \item F
  \end{itemize}
  \begin{itemize}
    \item A
    \item B
    \item C
    \item D
    \item E
    \item F
  \end{itemize}
  \begin{itemize}
    \item A
    \item B
    \item C
    \item D
    \item E
    \item F
  \end{itemize}
  \begin{itemize}
    \item A
    \item B
    \item C
    \item D
    \item E
    \item F
  \end{itemize}
  \begin{itemize}
    \item A
    \item B
    \item C
    \item D
    \item E
    \item F
  \end{itemize}
\end{itemize}
SUPERVISED LEARNING FOR BAYES NETS
Machine Learning

The **data** inspires the structures we want to predict.

---

**Inference** finds \{best structure, marginals, partition function\} for a new observation. *(Inference is usually called as a subroutine in learning)*

---

Our **model** defines a score for each structure and it also tells us what to optimize.

---

**Learning** tunes the parameters of the model.
Machine Learning

Data

(time flies like an arrow)

Inference

(Inference is usually called as a subroutine in learning)

Model

Objective

Learning

3 Alice saw Bob on a hill with a telescope

4 time flies like an arrow
Recipe for Closed-form MLE

1. Assume data was generated i.i.d. from some model (i.e. write the generative story)
   \( x^{(i)} \sim p(x|\theta) \)

2. Write log-likelihood
   \[ \ell(\theta) = \log p(x^{(1)}|\theta) + \ldots + \log p(x^{(N)}|\theta) \]

3. Compute partial derivatives (i.e. gradient)
   \[ \frac{\partial \ell(\theta)}{\partial \theta_1} = \ldots \]
   \[ \frac{\partial \ell(\theta)}{\partial \theta_2} = \ldots \]
   \[ \ldots \]
   \[ \frac{\partial \ell(\theta)}{\partial \theta_M} = \ldots \]

4. Set derivatives to zero and solve for \( \theta \)
   \[ \frac{\partial \ell(\theta)}{\partial \theta_m} = 0 \quad \text{for all} \quad m \in \{1, \ldots, M\} \]
   \( \theta_{\text{MLE}} = \text{solution to system of } M \text{ equations and } M \text{ variables} \)

5. Compute the second derivative and check that \( \ell(\theta) \) is concave down at \( \theta_{\text{MLE}} \)
Learning Fully Observed BNs

\[ p(X_1, X_2, X_3, X_4, X_5) = \]
\[ p(X_5|X_3)p(X_4|X_2, X_3) \]
\[ p(X_3)p(X_2|X_1)p(X_1) \]
Learning Fully Observed BNs

\[ p(X_1, X_2, X_3, X_4, X_5) = p(X_5|X_3)p(X_4|X_2, X_3) \]

\[ p(X_3)p(X_2|X_1)p(X_1) \]
Learning Fully Observed BNs

\[ p(X_1, X_2, X_3, X_4, X_5) = \]
\[ p(X_5 | X_3)p(X_4 | X_2, X_3) \]
\[ p(X_3)p(X_2 | X_1)p(X_1) \]

How do we learn these **conditional** and **marginal** distributions for a Bayes Net?
Learning Fully Observed BNs

Learning this fully observed Bayesian Network is equivalent to learning five (small / simple) independent networks from the same data.

\[
p(X_1, X_2, X_3, X_4, X_5) = p(X_5|X_3)p(X_4|X_2, X_3)p(X_3)p(X_2|X_1)p(X_1)
\]
Learning Fully Observed BNs

How do we learn these conditional and marginal distributions for a Bayes Net?

$$\theta^* = \arg\max_{\theta} \log p(X_1, X_2, X_3, X_4, X_5)$$

$$= \arg\max_{\theta} \log p(X_5|X_3, \theta_5) + \log p(X_4|X_2, X_3, \theta_4)$$

$$+ \log p(X_3|\theta_3) + \log p(X_2|X_1, \theta_2)$$

$$+ \log p(X_1|\theta_1)$$

$$\theta_1^* = \arg\max_{\theta_1} \log p(X_1|\theta_1)$$

$$\theta_2^* = \arg\max_{\theta_2} \log p(X_2|X_1, \theta_2)$$

$$\theta_3^* = \arg\max_{\theta_3} \log p(X_3|\theta_3)$$

$$\theta_4^* = \arg\max_{\theta_4} \log p(X_4|X_2, X_3, \theta_4)$$

$$\theta_5^* = \arg\max_{\theta_5} \log p(X_5|X_3, \theta_5)$$
Example: Tornado Alarms

1. Imagine that you work at the 911 call center in Dallas.
2. You receive six calls informing you that the Emergency Weather Sirens are going off.
3. What do you conclude?
Example: Tornado Alarms

1. Imagine that you work at the 911 call center in Dallas.
2. You receive six calls informing you that the Emergency Weather Sirens are going off.
3. What do you conclude?

Figure from https://www.nytimes.com/2017/04/08/us/dallas-emergency-sirens-hacking.html
Learning Fully Observed BNs

Ex: Toronto Alarms

\[ P(\cdots) = \prod_j p(x_j | \text{parent}(x_j)) \]

- \( H \sim \text{Bernoulli} (\theta) \)
- \( T \sim \text{Bernoulli} (\alpha) \)
- \( A \sim \text{Bernoulli} (\alpha_{H,T}) \)
- \( C \sim \text{Uniform}(1, \ldots, 63) + A \times \text{Uniform}(1, \ldots, 63) \)

No parameters

Data:

<table>
<thead>
<tr>
<th>T</th>
<th>H</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

MLEs in Closed Form

\[ \lambda(e, \theta, \alpha) = \frac{1}{N} \sum_{i=1}^{N} p(t^{(i)}, h^{(i)}, a^{(i)}, c^{(i)} | e, \theta, \alpha) \]

\[ \hat{\lambda}, \hat{\theta}, \hat{\alpha} = \arg\max \lambda(e, \theta, \alpha) \]

\[ \hat{\lambda} = \frac{1}{N} \sum_{i=1}^{N} \log p(t^{(i)} | e) = \frac{\#(T=t)}{N} \]

\[ \hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} \log p(h^{(i)} | e) = \frac{\#(H=h)}{N} \]

\[ \hat{\alpha} = \frac{1}{N} \sum_{i=1}^{N} \log p(a^{(i)} | e, h^{(i)}, \lambda) \]

\[ \hat{\alpha}_{t,h} = \frac{\#(A=t, T=t, H=h)}{\#(T=t, H=h)} \]
INFERENCE FOR BAYESIAN NETWORKS
A Few Problems for Bayes Nets

Suppose we already have the parameters of a Bayesian Network...

1. How do we compute the probability of a specific assignment to the variables?
   \[ P(T=t, H=h, A=a, C=c) = P(T=t) \cdot P(H=h) \cdot P(A=a \mid H=h, T=t) \cdot P(C=c \mid A=a) \]

2. How do we draw a sample from the joint distribution?
   \( t, h, a, c \sim P(T, H, A, C) \)  
   **Ancestral sampling**

3. How do we compute marginal probabilities?
   \( P(A) = \ldots \)

4. How do we draw samples from a conditional distribution?
   \( t, h, a \sim P(T, H, A \mid C = c) \)

5. How do we compute conditional marginal probabilities?
   \( P(H \mid C = c) = \ldots \)
Gibbs Sampling

\[ p(x) \]

\[ p(x_1 \mid x_2^{(t)}) \]

\( x \)

\[ x^{(t)} \]

\[ x^{(t+1)} \]
Gibbs Sampling

\[ p(x) \]

\[ p(x_2 | x_1^{(t+1)}) \]

\[ x^{(t+2)} \]

\[ x^{(t+1)} \]

\[ x^{(t)} \]
Gibbs Sampling

\[ p(x) \]

\[ x_1 \]

\[ x_2 \]

\( x(t) \)

\( x(t+1) \)

\( x(t+2) \)

\( x(t+3) \)

\( x(t+4) \)
Gibbs Sampling

**Question:**
How do we draw samples from a conditional distribution?

\[ y_1, y_2, \ldots, y_J \sim p(y_1, y_2, \ldots, y_J \mid x_1, x_2, \ldots, x_J) \]

**(Approximate) Solution:**
- Initialize \( y_1^{(0)}, y_2^{(0)}, \ldots, y_J^{(0)} \) to arbitrary values
- For \( t = 1, 2, \ldots \):
  - \( y_1^{(t+1)} \sim p(y_1 \mid y_2^{(t)}, \ldots, y_J^{(t)}, x_1, x_2, \ldots, x_J) \)
  - \( y_2^{(t+1)} \sim p(y_2 \mid y_1^{(t+1)}, y_3^{(t)}, \ldots, y_J^{(t)}, x_1, x_2, \ldots, x_J) \)
  - \( y_3^{(t+1)} \sim p(y_3 \mid y_1^{(t+1)}, y_2^{(t+1)}, y_4^{(t)}, \ldots, y_J^{(t)}, x_1, x_2, \ldots, x_J) \)
  - \( \ldots \)
  - \( y_J^{(t+1)} \sim p(y_J \mid y_1^{(t+1)}, y_2^{(t+1)}, \ldots, y_{J-1}^{(t+1)}, x_1, x_2, \ldots, x_J) \)

**Properties:**
- This will eventually yield samples from
  \[ p(y_1, y_2, \ldots, y_J \mid x_1, x_2, \ldots, x_J) \]
- But it might take a long time -- just like other Markov Chain Monte Carlo methods
Gibbs Sampling

\[ p(x_6 | x_1, x_2, \ldots, x_{13}) = p(x_6 | x_3, x_4, x_5, x_8, x_9, x_{10}) \]

Full conditionals only need to condition on the Markov boundary

- Must be “easy” to sample from conditionals
- Many conditionals are log-concave and are amenable to adaptive rejection sampling
Learning Objectives

Bayesian Networks

You should be able to...

1. Identify the conditional independence assumptions given by a generative story or a specification of a joint distribution
2. Draw a Bayesian network given a set of conditional independence assumptions
3. Define the joint distribution specified by a Bayesian network
4. Use domain knowledge to construct a (simple) Bayesian network for a real-world modeling problem
5. Depict familiar models as Bayesian networks
6. Use d-separation to prove the existence of conditional independencies in a Bayesian network
7. Employ a Markov boundary to identify conditional independence assumptions of a graphical model
8. Develop a supervised learning algorithm for a Bayesian network
9. Use samples from a joint distribution to compute marginal probabilities
10. Sample from the joint distribution specified by a generative story
11. Implement a Gibbs sampler for a Bayesian network

...eh...
LEARNING PARADIGMS
## Learning Paradigms

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>$\mathcal{D} = {\mathbf{x}^{(i)}, y^{(i)}}_{i=1}^{N}$ \quad $\mathbf{x} \sim p^<em>(\cdot)$ and $y = c^</em>(\cdot)$</td>
</tr>
<tr>
<td>$\rightarrow$ Regression</td>
<td>$y^{(i)} \in \mathbb{R}$</td>
</tr>
<tr>
<td>$\rightarrow$ Classification</td>
<td>$y^{(i)} \in {1, \ldots, K}$</td>
</tr>
<tr>
<td>$\rightarrow$ Binary classification</td>
<td>$y^{(i)} \in {+1, -1}$</td>
</tr>
<tr>
<td>$\rightarrow$ Structured Prediction</td>
<td>$y^{(i)}$ is a vector</td>
</tr>
</tbody>
</table>
## Learning Paradigms

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Supervised</strong></td>
<td>$D = {x^{(i)}, y^{(i)}}_{i=1}^N$ $\mathbf{x} \sim p^<em>(\cdot)$ and $y = c^</em>(\cdot)$</td>
</tr>
<tr>
<td>$\hookrightarrow$ Regression</td>
<td>$y^{(i)} \in \mathbb{R}$</td>
</tr>
<tr>
<td>$\hookrightarrow$ Classification</td>
<td>$y^{(i)} \in {1, \ldots, K}$</td>
</tr>
<tr>
<td>$\hookrightarrow$ Binary classification</td>
<td>$y^{(i)} \in {+1, -1}$</td>
</tr>
<tr>
<td>$\hookrightarrow$ Structured Prediction</td>
<td>$y^{(i)}$ is a vector</td>
</tr>
<tr>
<td><strong>Unsupervised</strong></td>
<td>$D = {x^{(i)}}_{i=1}^N$ $\mathbf{x} \sim p^*(\cdot)$</td>
</tr>
</tbody>
</table>
# Learning Paradigms

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>$\mathcal{D} = { \mathbf{x}^{(i)}, y^{(i)} }_{i=1}^{N}$ $\mathbf{x} \sim p^<em>(\cdot)$ and $y = c^</em>(\cdot)$</td>
</tr>
<tr>
<td>←→ Regression</td>
<td>$y^{(i)} \in \mathbb{R}$</td>
</tr>
<tr>
<td>←→ Classification</td>
<td>$y^{(i)} \in {1, \ldots, K}$</td>
</tr>
<tr>
<td>←→ Binary classification</td>
<td>$y^{(i)} \in {+1, -1}$</td>
</tr>
<tr>
<td>←→ Structured Prediction</td>
<td>$\mathbf{y}^{(i)}$ is a vector</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>$\mathcal{D} = { \mathbf{x}^{(i)} }_{i=1}^{N}$ $\mathbf{x} \sim p^*(\cdot)$</td>
</tr>
<tr>
<td>Semi-supervised</td>
<td>$\mathcal{D} = { \mathbf{x}^{(i)}, y^{(i)} }<em>{i=1}^{N_1} \cup { \mathbf{x}^{(j)} }</em>{j=1}^{N_2}$</td>
</tr>
</tbody>
</table>
# Learning Paradigms

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>$\mathcal{D} = {\mathbf{x}^{(i)}, y^{(i)}}_{i=1}^N$  (\mathbf{x} \sim p^<em>(\cdot)) and (y = c^</em>(\cdot))</td>
</tr>
<tr>
<td>Regression</td>
<td>(y^{(i)} \in \mathbb{R})</td>
</tr>
<tr>
<td>Classification</td>
<td>(y^{(i)} \in {1, \ldots, K})</td>
</tr>
<tr>
<td>Binary classification</td>
<td>(y^{(i)} \in {+1, -1})</td>
</tr>
<tr>
<td>Structured Prediction</td>
<td>(y^{(i)}) is a vector</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>$\mathcal{D} = {\mathbf{x}^{(i)}}_{i=1}^N$  (\mathbf{x} \sim p^*(\cdot))</td>
</tr>
<tr>
<td>Semi-supervised</td>
<td>$\mathcal{D} = {\mathbf{x}^{(i)}, y^{(i)}}<em>{i=1}^{N_1} \cup {\mathbf{x}^{(j)}}</em>{j=1}^{N_2}$</td>
</tr>
<tr>
<td>Online</td>
<td>$\mathcal{D} = {(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}), (\mathbf{x}^{(3)}, y^{(3)}), \ldots}$</td>
</tr>
</tbody>
</table>
# Learning Paradigms

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>$\mathcal{D} = {x^{(i)}, y^{(i)}}_{i=1}^N$  \hspace{1cm} $x \sim p^<em>(\cdot)$ and $y = c^</em>(\cdot)$</td>
</tr>
<tr>
<td>Regression</td>
<td>$y^{(i)} \in \mathbb{R}$</td>
</tr>
<tr>
<td>Classification</td>
<td>$y^{(i)} \in {1, \ldots, K}$</td>
</tr>
<tr>
<td>Binary classification</td>
<td>$y^{(i)} \in {+1, -1}$</td>
</tr>
<tr>
<td>Structured Prediction</td>
<td>$y^{(i)}$ is a vector</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>$\mathcal{D} = {x^{(i)}}_{i=1}^N$  \hspace{1cm} $x \sim p^*(\cdot)$</td>
</tr>
<tr>
<td>Semi-supervised</td>
<td>$\mathcal{D} = {x^{(i)}, y^{(i)}}<em>{i=1}^{N_1} \cup {x^{(j)}}</em>{j=1}^{N_2}$</td>
</tr>
<tr>
<td>Online</td>
<td>$\mathcal{D} = {(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \ldots}$</td>
</tr>
<tr>
<td>Active Learning</td>
<td>$\mathcal{D} = {x^{(i)}}_{i=1}^N$ and can query $y^{(i)} = c^*(\cdot)$ at a cost</td>
</tr>
</tbody>
</table>
## Learning Paradigms

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>( \mathcal{D} = {x^{(i)}, y^{(i)}}_{i=1}^N \quad x \sim p^<em>(\cdot) \text{ and } y = c^</em>(\cdot) )</td>
</tr>
<tr>
<td>Regression</td>
<td>( y^{(i)} \in \mathbb{R} )</td>
</tr>
<tr>
<td>Classification</td>
<td>( y^{(i)} \in {1, \ldots, K} )</td>
</tr>
<tr>
<td>Binary classification</td>
<td>( y^{(i)} \in {+1, -1} )</td>
</tr>
<tr>
<td>Structured Prediction</td>
<td>( y^{(i)} ) is a vector</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>( \mathcal{D} = {x^{(i)}}_{i=1}^N \quad x \sim p^*(\cdot) )</td>
</tr>
<tr>
<td>Semi-supervised</td>
<td>( \mathcal{D} = {x^{(i)}, y^{(i)}}<em>{i=1}^{N_1} \cup {x^{(j)}}</em>{j=1}^{N_2} )</td>
</tr>
<tr>
<td>Online</td>
<td>( \mathcal{D} = {(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \ldots} )</td>
</tr>
<tr>
<td>Active Learning</td>
<td>( \mathcal{D} = {x^{(i)}}_{i=1}^N ) \text{ and can query } y^{(i)} = c^*(\cdot) \text{ at a cost}</td>
</tr>
<tr>
<td>Imitation Learning</td>
<td>( \mathcal{D} = {(s^{(1)}, a^{(1)}), (s^{(2)}, a^{(2)}), \ldots} )</td>
</tr>
</tbody>
</table>
## Learning Paradigms

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>$\mathcal{D} = {\mathbf{x}^{(i)}, y^{(i)}}_{i=1}^N$ $\mathbf{x} \sim p^<em>(\cdot)$ and $y = c^</em>(\cdot)$</td>
</tr>
<tr>
<td>$\rightarrow$ Regression</td>
<td>$y^{(i)} \in \mathbb{R}$</td>
</tr>
<tr>
<td>$\rightarrow$ Classification</td>
<td>$y^{(i)} \in {1, \ldots, K}$</td>
</tr>
<tr>
<td>$\rightarrow$ Binary classification</td>
<td>$y^{(i)} \in {+1, -1}$</td>
</tr>
<tr>
<td>$\rightarrow$ Structured Prediction</td>
<td>$y^{(i)}$ is a vector</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>$\mathcal{D} = {\mathbf{x}^{(i)}}_{i=1}^N$ $\mathbf{x} \sim p^*(\cdot)$</td>
</tr>
<tr>
<td>Semi-supervised</td>
<td>$\mathcal{D} = {\mathbf{x}^{(i)}, y^{(i)}}<em>{i=1}^N \cup {\mathbf{x}^{(j)}}</em>{j=1}^{N_2}$</td>
</tr>
<tr>
<td>Online</td>
<td>$\mathcal{D} = {(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}), (\mathbf{x}^{(3)}, y^{(3)}), \ldots}$</td>
</tr>
<tr>
<td>Active Learning</td>
<td>$\mathcal{D} = {\mathbf{x}^{(i)}}_{i=1}^N$ and can query $y^{(i)} = c^*(\cdot)$ at a cost</td>
</tr>
<tr>
<td>Imitation Learning</td>
<td>$\mathcal{D} = {(s^{(1)}, a^{(1)}), (s^{(2)}, a^{(2)}), \ldots}$</td>
</tr>
<tr>
<td>Reinforcement Learning</td>
<td>$\mathcal{D} = {(s^{(1)}, a^{(1)}, r^{(1)}), (s^{(2)}, a^{(2)}, r^{(2)}), \ldots}$</td>
</tr>
</tbody>
</table>
REINFORCEMENT LEARNING
Reinforcement Learning

Source: https://www.xkcd.com/242/
RL: Examples

Source: https://twitter.com/alphagomovie

Source: https://techobserver.net/2019/06/argo-ai-self-driving-car-research-center/  
Source: https://www.wired.com/2012/02/high-speed-trading/
AlphaGo

Source: https://www.youtube.com/watch?v=WXuK6gekU1Y&ab_channel=DeepMind
History of Reinforcement Learning

• Roots in the **psychology of animal learning** (Thorndike, 1911).

• Another independent thread was the problem of **optimal control**, and its solution using **dynamic programming** (Bellman, 1957).

• Idea of **temporal difference** learning (on-line method), e.g., playing board games (Samuel, 1959).

• A major breakthrough was the discovery of **Q-learning** (Watkins, 1989).
What is special about RL?

• RL is learning how to map states to actions, so as to maximize a numerical reward over time.

• Unlike other forms of learning, it is a multistage decision-making process (often Markovian).

• An RL agent must learn by trial-and-error. (Not entirely supervised, but interactive)

• Actions may affect not only the immediate reward but also subsequent rewards (Delayed effect).
Elements of RL

• A policy
  - A map from state space to action space.
  - May be stochastic.

• A reward function
  - It maps each state (or, state-action pair) to a real number, called reward.

• A value function
  - Value of a state (or, state-action pair) is the total expected reward, starting from that state (or, state-action pair).
Example: Robot in a Room

- reward +1 at [4,3], -1 at [4,2]
- reward -0.04 for each step

actions: UP, DOWN, LEFT, RIGHT

UP
80% move UP
10% move LEFT
10% move RIGHT
Example: Robot in a Room

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>+1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>▼</td>
<td>▼</td>
</tr>
<tr>
<td></td>
<td></td>
<td>▼</td>
<td>▼</td>
</tr>
<tr>
<td></td>
<td></td>
<td>▼</td>
<td>▼</td>
</tr>
<tr>
<td></td>
<td></td>
<td>▼</td>
<td>▼</td>
</tr>
</tbody>
</table>

Question: Is this policy optimal: yes or no? Briefly justify your answer.

Answer: (Hint: both yes and no are acceptable answers, I’m interested in your justification.)
Example: Robot in a Room

• Reward for each step -2
Example: Robot in a Room

• Reward for each step: -0.1
The Precise Goal

• To find a policy that maximizes the Value function.
  – transitions and rewards usually not available

• There are different approaches to achieve this goal in various situations.

  • **Value iteration** and **Policy iteration** are two more classic approaches to this problem. But essentially both are **dynamic programming**.

  • **Q-learning** is a more recent approaches to this problem. Essentially it is a **temporal-difference method**.
MARKOV DECISION PROCESSES
RL: Components

From the Environment (i.e. the MDP)
- State space, $S$
- Action space, $A$
- Reward function, $R(s, a), R : S \times A \rightarrow \mathbb{R}$
- Transition probabilities, $p(s' \mid s, a)$
  - Deterministic transitions:
    \[
    p(s' \mid s, a) = \begin{cases} 
    1 & \text{if } \delta(s, a) = s' \\
    0 & \text{otherwise}
    \end{cases}
    \]
    where $\delta(s, a)$ is a transition function

From the Model
- Policy, $\pi : S \rightarrow A$
- Value function, $V^\pi : S \rightarrow \mathbb{R}$
  - Measures the expected total payoff of starting in some state $s$ and executing policy $\pi$

Markov Assumption
\[
p(s_{t+1} \mid s_t, a_t, \ldots, s_1, a_1) = p(s_{t+1} \mid s_t, a_t)
\]
Markov Decision Process (MDP)

• For **supervised learning** the **PAC learning framework** provided assumptions about where our data came from:

\[ x \sim p^*(\cdot) \text{ and } y = c^*(\cdot) \]

• For **reinforcement learning** we assume our data comes from a **Markov decision process (MDP)**
Markov Decision Processes (MDP)

In RL, the source of our data is an MDP:

1. Start in some initial state $s_0 \in S$
2. For time step $t$:
   1. Agent observes state $s_t \in S$
   2. Agent takes action $a_t \in A$ where $a_t = \pi(s_t)$
   3. Agent receives reward $r_t \in \mathbb{R}$ where $r_t = R(s_t, a_t)$
   4. Agent transitions to state $s_{t+1} \in S$ where $s_{t+1} \sim p(s' | s_t, a_t)$
3. Total reward is $\sum_{t=0}^{\infty} \gamma^t r_t$
   - The value $\gamma$ is the “discount factor”, a hyperparameter $0 < \gamma < 1$

• Makes the same Markov assumption we used for HMMs! The next state only depends on the current state and action.
• Def.: we execute a policy $\pi$ by taking action $a = \pi(s)$ when in state $s$