CS 232: Artificial Intelligence

Spring 2024

Prof. Carolyn Anderson Wellesley College

Reminders

- I have help hours today from 3:30-4:30 and on Monday from 4-5:15
- Midterm is 1 week away (no HW until then)
- Bring midterm questions to class on Tuesday, we'll have a brief review
- Read YLLATAILY Chapters 5-6 for next Tuesday

How To Explore

Passive Reinforcement Learning

Take a fixed policy and follow it

Random Exploration

Pick actions randomly



Balancing Exploration and Exploitation

 Pick actions randomly initially, but gradually switch to mostly doing actions that you have already found to be valuable

How To Evaluate

In Direction Evaluation, we repeatedly follow the same policy and observe its rewards.

Direct Evaluation:

- Act according to policy π
- In each state, write down the sum of discounted rewards
- Average samples

Example: Direct Evaluation



Example: Direct Evaluation













Example: Direct Evaluation



Weaknesses of Direct Evaluation

Direct Evaluation works.

But, we had to learn each state separately.

That isn't very efficient!

Output Values



If B and E both go to C under this policy, how can their values be different? Temporal Difference Learning

Temporal Difference Learning

Main Idea: learn from every experience

TD Learning

Update V(s) every time we make a transition T(s,a,s',r)

& · learning

Most likely outcomes (s') will contribute more updates

Sample of V(s): Update to V(s): Same update:

1 etc

"How willing one

me to believe

von enice

TD: A Moving Average

In TD Learning, we average our observations in a way that weights more recent samples highly.

Over time, we forget our initial estimates (which probably weren't very good!)

$$\bar{x}_n = \frac{x_n + (1 - \alpha) \cdot x_{n-1} + (1 - \alpha)^2 \cdot x_{n-2} + \dots}{1 + (1 - \alpha) + (1 - \alpha)^2 + \dots}$$

We decrease the **learning rate** α over time.











 $\alpha = 1/2$

Q-Learning

In practice, we want to know the value of state-action pairs, not just states.

This version of TD Learning is called **Q-Learning**, because the value of a state-action pair is called a **q-value**.





Demo of Q-Learning



Regret

Even if our Reinforcement Learning agent learns an optimal policy eventually, it will still take sub-par actions along the way.

Regret is the total cost of all of those mistakes: the differences between optimal expected rewards and the agent's actual rewards.

The best RL agent is one that **minimizes regret**: learns optimally!



Semester Road Map

Almost all AI tasks can be grouped into one of three main categories:

- Search
- Classification
- Generation

Classification Methods

Classification Methods: Supervised Machine Learning

Lots of kinds!

- Naïve Bayes
- Logistic regression
- Neural networks
- k-Nearest Neighbors
- random forests

0

Classification Methods: Supervised Machine Learning

Input:

- \circ a input d
- a fixed set of classes $C = \{c_1, c_2, \dots, c_j\}$
- A training set of *m* hand-labeled examples
 (d₁, c₁), (dm, cm)

Output:

• a learned classifier $y: d \rightarrow C$

Components of a probabilistic machine learning classifier

Given *m* input/output pairs $(x^{(i)}, y^{(i)})$:

- A feature representation of the input For each input x', a vector of features [x, x. xn] Feature j for x' as xj or xj
 A classification function that fathes x and compute ŷ, the estimated class later for x by estimating p(y|x)
- 3. An objective function for learning, like **cross-entropy loss**
- An algorithm for optimizing the objective function: stochastic gradient descent.

Break: Classifiers and Features

Classifiers & Features

Logistic Regression Classifiers

Is this spam?

Subject: Important notice!

From: Stanford University <newsforum@stanford.edu>

- Date: October 28, 2011 12:34:16 PM PDT
- To: undisclosed-recipients:;

Greats News!

You can now access the latest news by using the link below to login to Stanford University News Forum.

http://www.123contactform.com/contact-form-StanfordNew1-236335.html

Click on the above link to login for more information about this new exciting forum. You can also copy the above link to your browser bar and login for more information about the new services.

© Stanford University. All Rights Reserved.

Who wrote which Federalist papers?

Anonymous essays try to convince New York to ratify U.S Constitution. Authorship of 12 of the letters in dispute.

Solved by Mosteller and Wallace (1963) using Bayesian methods

...



What is the subject of this research article?

MEDLINE Article

	aciercedirect.com Billionari Particular	Brain Cognition
Syntactic frame and verb bias in of undergoer-su	aphasia: Plausibi ibject sentences	lity judgments
Sasama Gahi," Lise Mann," Gali Ramshayar," Danial S. Jorafay," Blacketh Eddar," Molly Rawaya, "and L. Falanai Auday " anarchicany statistic as a set " anarchicany change and a " anarchicany change as an " anarchicany change as an		
difficulties enablings on these of passes and man. Using a place professional approximation for the structure of a structure over an enable of exploration of passion from other the structure approach encourses. Passions of passion from other area significant approach encourses. Passions of passion from other area significant approach encourses. Passions of passion from other area significant approach encourses.	alting julgers and a disc by based has, in the failers is also for particular failers	la a stal project of option options of provide the for a stategy provide pro- location. Note provide pro-
effective evolution or chern of scalar entries. Using a fee performant optimization of scalar entries that are perform that entries for optime. We share that this offset is entried and space is ensured to the first of space is the scalar entries of scalar entries. The independence of the scalar entries of the is the scalar transmission is the first of the scalar entries of the scalar entries of the independence of the scalar entries of the scalar entries. All optimized as	and by Segment and a second by here the La, for Failler 1 and the particular second a particular for particular of an and a second segment and the second se	na antaŭ prograf (j. 1996) 1996 - Partin Prograf (j. 1996) 1997 - Partin Prograf (j. 1997) 1997 - Partin Prograf (j. 1997) 199
Charlier and San	alting pagasan and, walker in the face of a late of a standard in age of the page of the standard in age of the page of the standard in age of the page of the standard in age of the standard of the office of the standard o	
Display the end of the second seco	And the programment was a section of the state over the section of	the a return proof of a black or particular or proof of a black in a case a proof of a black in a case a proof of a black in a case of the star of a black in a black of the star of a black of the star of the star of a black of the star of the star of a black of the star
Elitation of page at the of page and page at the page of page of page page page at the page at the page at the page page page at the page page at the page at the page page at the page at the page page page at the page page at the page at the page page page at the page page at the page at the page page page at the page page at the page at the page page at the page page at the page page at the page page page at the page page at the page page at the page page page at the page page at the page page at the page page page at the page page at the page page at the page page at the page page page page page page at the page page page page page page at the page page page page page page page pag	and a polycer with an extension of the frame way with the extension of the second way in the framework of the second second second second and the framework of the second second second second second second second second second second second second the second second second second second second second second second second second second second second second second second second	The control prove of columns of the control prove of columns from a net registerior to given the control term of the column interview of the column of the proves of the column of the proves of the column of the column of the control term of the column of the control term of the column of the control term of the column of the column provided with the term of the column of the column of of the term of the column of the
Experimental program of these and these first program of the pr	The property of the second sec	The constraint prove of columns in a constraint property having the a constraint property having the constraint provide a constraint of the constraint of th
Experimental estimation of address and the first product of the second secon	and a paper of the strength	The contrast prove of elements or granteness granteness provides the the second system is a place the second system is a place to extra the system is a place to extra the system is a second system of the system is a second system is a system is a function of the system is a second provides to obtain any granteness provides to obtain a second system is a second on Regression is a second as a second system. Under the is a second system is a second in the second system is a second system is a second system is a second system is a second system is a second system is a second system is a second system is a second in the second system is a second system is a second system is a second system is a second in the second system is a second system is a second system is a second system is a second in the second system is a second system is a second system is a second system is a second in the second system is a second system is a second system is a second system is a second in the second system is a second system is a second system is a second system is a second in the second system is a second system is a second in the second system is a second system is a second in the second system is a second system is a second in the second system is a second system is a second in the second system is a second system is a second in the second system is a second system is a second in the second system is a second system is a second system is a second in the second system is a second system is a second system is a second in the second system is a second system is a second system is a second in the second system is a second
Experimental product of a set of the set	and the program was a set of the second of the second second sec	The entropy of entropy of entropy of the entropy of
Experimental processing of the second sec	and by pages in the set of the second set of	The entropy of the description of a power of the description of a power of the description of a power of the description of the
How see the second s	The property of the second sec	The ansatz prove of characteristic of the second se
Construction of the second	and by papers when the other section is a transmission of the section of the s	The answer of the second secon
However, and the second sec	and by particular that a start of the start	The answer of the second secon
Construction of the second	and the property with a loss of the property of the second	The second secon
An	and the property with a loss of the property o	The answer of the second secon

Antogonists and Inhibitors Blood Supply Chemistry Drug Therapy Embryology Epidemiology

Text Classification: definition

Input:

- ° a document d
- a fixed set of classes $C = \xi_{c_1, c_2, \dots, c_j}$

Output: a predicted class $\dot{y} \in C$

Binary Classification in Logistic Regression

Given a series of input/output pairs: (x^{i}, y^{i})

For each observation x⁽ⁱ⁾

"We represent x⁽ⁱ⁾ by a forture vector $[x_1, x_2, ..., x_n]$

• We compute an output

ŷ' E {0,13

Features in logistic regression

For feature x_i , weight w_i tells is how important is x_i

- $x_i =$ "review contains 'marvelous'": $\omega_i = +10$
- $x_i =$ "review contains 'awful": $w_i = -5$
- x_k = "review contains 'disappoint'": wk - 2

Logistic Regression for one observation x

Input observation: $x = (x_1, x_2, ..., x_n]$ Weights (one per feature): $W = [w_1, w_2, ..., m]$ $\theta = [0, ..., \theta_2, ..., \theta_n]$ Output: a predicted class $\hat{y} \notin \hat{\xi}_{0,1}$ Ney Pos

How to do classification

For each feature x_i , weight w_i tells us importance of x_i

We'll sum up all the weighted features and the bias:

$$z = \begin{pmatrix} n \\ \sum_{i=1}^{n} N_i x_i \end{pmatrix} + b$$

$$z = W X + b$$

$$f \qquad \text{intercept}$$

$$dye \qquad iF z \ge 0 : we say 1 (+)$$

$$Otherwise : we dole 0 (-)$$

But we want a probabilistic classifier

We need to formalize "sum is high". We want a model that can tell us:

The very useful sigmoid or logistic function

