CPSC 340: Machine Learning and Data Mining

Ranking Fall 2016

Admin

- Assignment 5:
 - 2 late days to hand in Wednesday, 3 for Friday.
- Assignment 6:
 - Due Friday, 1 late day to hand in next Monday, etc.
- Final:
 - December 12 (8:30am HEBB 100)
 - Covers Assignments 1-6.
 - List of topics posted.
 - Final from last year will be posted Friday.
 - Closed-book, cheat sheet: 4-pages each double-sided.

Ranking

- The ranking problem:
 - Input: a set of objects and some information about "ordering".



- Output: an ordering of the objects.



Ranking

- The ranking problem:
 - Input: a set of objects and some measure of relative "ordering".
 - Output: an ordering of the objects.
- Examples:
 - Country comparisons (Global Hunger Index).
 - Academic journals (Impact factor).
 - Sports/gaming (Elo and TrueSkill).
 - Internet search engines.
- Large, diverse, and well-studied topic.
 - We focus on learning to rank.

http://chess.stackexchange.com/questions/2550/whats-the-average-elo-rating-whats-the-average-uscf-rating https://commons.wikimedia.org/wiki/File:GHI_2008_map.jpg



About 658,000,000 results (0.37 seconds)

Ranking - Wikipedia, the free encyclopedia https://en.wikipedia.org/wiki/Ranking •

A ranking is a relationship between a set of items such that, for any two items, the first is either 'ranked higher than', 'ranked lower than' or 'ranked equal to' the ... Strategies for assigning rankings - Ranking in statistics - Examples of ranking You visited this page on 16/11/15.

University Rankings | Top Universities www.topuniversities.com/university-rankings •

QS University Rankings: Arab Region 2015. ... Compare the world's highest-performing universities with the latest edition of the QS World University Rankings®, and explore the leading universities in different world regions and in specific subject areas. ... Discover the world's top ...

QS World University Rankings - QS University Rankings: Asia - QS Top 50 Under 50 $\,$

QS World University Rankings® 2015/16 | Top Universities www.topuniversities.com > Rankings > World University Rankings ▼

Welcome to the QS World University Rankings® 2015/16. Use the interactive ranking table to explore the world's top universities, with options to sort the results ...

Ranking Web of Universities www.webometrics.info/ A directory of world universities ranked according their presence on the Web





Supervised Ranking

- Ranking based on supervised learning:
 - We have features x_i for each object 'i', and "information" about labels y_i .
- Forms the "information" can take:
 - 1. Item relevance:
 - Explicit numerical "scores" y_i.
 - 2. Pairwise preference:
 - Pairs 'i' and 'j' where we know $y_i > y_j$.
 - But we don't know the "score" of any items.
 - 3. Total/partial ordering:
 - Larger sets of items where we know $y_i > y_j > y_k > y_m > ...$

Supervised Ranking with Item Relevance

- With item relevance we have explicit "score" y_i for each object 'i'.
- We can rank with regression:

Find 'w' minimizing
$$f(w) = \frac{1}{2} \sum_{i=1}^{2} (w^{7}x_{i} - y_{i})^{2}$$

• Compute score of new object 'i' based on its features 'x_i'.

$$\gamma_i = w^7 x_i$$

• If scores are ordinal {1,2,3,..,k}, can use ordinal logistic regression.

Supervised Ranking with Query

- Common variation on ranking includes query 'q'.
 - E.g., for web search it could the keywords.
- Can adapt item relevance to this setting:
 - Measure features x_{iq} of object/query combination.
 - Item relevance y_{ig} gives "score" of object/query combination.

- Unfortunately, item relevance may be hard to get:
 - Active human effort to produce meaningful labels across queries/objects.
 - How do you compare 'CPSC 340' to 'shoe' or 'moon' to 'Tuesday' on same scale?

Supervised Ranking with Pairwise Preferences

- More realistic is pairwise preferences:
 - We aren't given any explicit y_i values.
 - Instead we're given list of objects (i,j) where $y_i > y_i$.
 - E.g., which one looks more like 'smoke':
 - Much easier than asking artist for score.



Can we design a loss function with this label information?

https://circle.ubc.ca/bitstream/handle/2429/30519/ubc_2011_spring_brochu_eric.pdf?sequence=3

- We've seen that loss functions can come from probabilities:
 Gaussian => squared loss, Laplace => absolute loss, sigmoid => logistic.
- Most other loss functions can be derived from probability ratios.
 - Example: sigmoid => hinge.

$$p(y_i | x_{i,w}) = \frac{1}{1 + exp(-y_i w^T x_i)} = \frac{exp(\frac{1}{2}y_i w^T x_i)}{exp(\frac{1}{2}y_i w^T x_i) + exp(-\frac{1}{2}y_i w^T x_i)} \propto exp(\frac{1}{2}y_i w^T x_i)$$
Same normalizing constant
for $y_i = +1$ and $y_i = -1$

- We've seen that loss functions can come from probabilities:
 Gaussian => squared loss, Laplace => absolute loss, sigmoid => logistic.
- Most other loss functions can be derived from probability ratios.
 Example: sigmoid => hinge.

 $p(y_i | x_{ij}w) \propto exp(\frac{1}{2} y_i w^T x_i)$ $T_0 \ classify \ y_i \ correctly, \ it's \ sufficient \ to \ have \ \frac{p(y_i | x_{ij}w)}{p(-y_i | x_{ij}w)} \not\supset \beta \ for \ some \ \beta' \ge 1$ $Notice \ theil \ normalizing \ constant \ dorsn't \ matter:$ $\frac{exp(\frac{1}{2} y_i w^T x_i)}{exp(-\frac{1}{2} y_i w^T x_i)} \not\supseteq \beta$

- We've seen that loss functions can come from probabilities:
 Gaussian => squared loss, Laplace => absolute loss, sigmoid => logistic.
- Most other loss functions can be derived from probability ratios.
 Example: sigmoid => hinge.

```
p(y_{i} | x_{i}, w) \propto exp(\frac{1}{2} y_{i}, w^{T} x_{i})
We neel: \underbrace{exp(\frac{1}{2} y_{i}, w^{T} x_{i})}_{exp(-\frac{1}{2} y_{i}, w^{T} x_{i})} \not\ni \beta
Take l_{\underline{oq}}:
\int d_{\underline{oq}} \left( \frac{e x p(\frac{1}{2} y_{i}, w^{T} x_{i})}{e x p(-\frac{1}{2} y_{i}, w^{T} x_{i})} \right) \not\supseteq l_{\underline{oq}}(\beta) \iff \frac{1}{2} y_{i} w^{T} x_{i} + \frac{1}{2} y_{i} w^{T} x_{i} \not\supseteq l_{\underline{oq}}(\beta)
```

- We've seen that loss functions can come from probabilities:
 Gaussian => squared loss, Laplace => absolute loss, sigmoid => logistic.
- Most other loss functions can be derived from probability ratios.

– Example: sigmoid => hinge.

 $p(y_i | x_{i,w}) \propto exp(\frac{1}{2} y_{i,w}^{T} x_{i})$ We need: $exp(\frac{1}{2} y_{i,w}^{T} x_{i}) \ge \beta$ $exp(-\frac{1}{2} y_{i,w}^{T} y_{i}) \ge \beta$ Or equivalently:

$$y_i w x_i \ge 1$$
 (for $\beta = exp(1)$)

Define a loss function by amount of constraint violation: max 20, 1 - yiw xi} when 1-yiw xi > 0 when 1-yiw xi > 0 We get SUMs by looking at regularized average loss: f(w) = Emax 20,1-yiw xi > + = 1/w/2

- General approach for defining losses using probability ratios:
 - 1. Define constraint based on probability ratios.
 - 2. Minimize violation of logarithm of constraint.
- Example: softmax => multi-class SVMs.

Assume:
$$p(y_i = c \mid x_{i,1}w) \propto exp(w_c^{T}x_i^{T})$$

Want: $p(y_i \mid x_{i,1}w) \Rightarrow \beta$ for all c'
 $p(y_i = c' \mid x_{i,1}w) \Rightarrow \beta$ for all c'
 $p(y_i = c' \mid x_{i,1}w) \Rightarrow \beta$ for all c'
and some $\beta \neq 1$
for $\beta = exp(1)$ equivalent to
 $W_{y_i}^{T}x_i = W_c^{T}x_i \Rightarrow 1$
for all c' $\neq y_i$
 $w_{y_i}^{T}x_i = W_c^{T}x_i \Rightarrow 1$
 $for all c' \neq y_i$
 $w_{y_i}^{T}x_i = W_c^{T}x_i \Rightarrow 1$
 $for all c' \neq y_i$

Supervised Ranking with Pairwise Preferences

- Ranking with pairwise preferences:
 - We aren't given any explicit y_i values.
 - Instead we're given list of objects (i,j) where $y_i > y_j$.

Assume
$$p(y_i | X_i w) \propto exp(w^T x_i)$$
 is probability that object 'i' has highest rank.
Want: $\frac{p(y_i | X_i w)}{p(y_j | X_i w)} \not\equiv \beta$ for all preferences (i, j)
For $\beta = exp(i)$ equivalent to
 $w^T x_i - w^T x_j \not\equiv 1$
For preferences (i, j)
This approach can also be used to define losses
for preferences (i, j)
This approach can also be used to define losses
for total/partial orderings. (but this information is hardlo
get

(pause)

Web Search before Google

Multi Search university	Search <u>Next! [national parks]</u>	<u></u>
	/	
Query: university 11 Results Returned Showing Results From 0 to 10	Optical Physics at the University of Oregon Oregon Center for Optics in Science and Technology. Department of Physics, University of Oregon, Eugene OR 97403. Research Groups Carmichael Group <u>http://typtich.ucongrun.ed.tt/</u> -site 1K - 16 Der 96	<u>н</u> г:
Stanford University Homepage http://www.stanford.etu/ 74.79% & - 459/1933 - 0100397 Stanford University, Portfolio Collection http://www.stanford.etu/home/alministration/portfolio.html 65.78% & - 459/1933 - 0100397 University of Illinois at Urbana-Champaign http://www.sias.abu/ 73.26% ISE - 12000/96 - 0100097 University	Carnegie Mellon University - Campus Networking Departments. Data Communications. Data Communications is responsible for installing and maintaining all on campus networking equipment and all of http://www.net.comm.edu/ bittp://www.net.comm.edu/ computer Science Group Home Page Computer Science Group. Wesleyan University. Welcome to the hom page of the Computer Science Group at Wesleyan University. We are administratively within. http://www.comment.com/edu/ http://www.comment.com/edu/	Due t
Indiana University http://www.indiana.etu/ 68.38% Ze - 0902099 - 02025907 University of California, Irvine http://www.uci.etu/ 68.07% Ze - 2202095 - 02023907 University of Minnesota http://www.uci.etu/	Keio University Shonan Fujisawa Campus (SFC) B\$3\$N%2/EFnF4B1%-%0%8%0%9 B(SFC) \$B\$N (BWWW \$I \$BCm0U=q\$- (B \$B\$nFI\$s\$G\$/\$@\$5\$\$\$# (B. Nihongo English SFC \$B>pJs (B. [\$B%8%0%8%%%%9%?]* http://www.sfc.keio.ac.fb/~size \$K - 5 Feb \$? School of Chemistry, University of Sydney The School of Chemistry. School of Chemistry, University of Sydney NSW 2006 Australia International Phone: +61-2-9351-4504 Fex:	896 h.
67.05% OR - 22/28/98 - 02/08/987 Iowa State University Homepage http://www.iotate.adu/ 66.06% SR - 22/28/98 - 02/08/987 The University of Michigan http://www.unich.edu/ 66.35% IR - 425/29/98 - 02/08/97	+61-2-9351-3329 Australia. <u>http://www.w.cheun.su.cu.au/</u> -size 4K - 25 Feb 97 Mankato State University The Campus Athletics, Campus Tour, Bookstore, Maps, Current Events Admission & Registration Admissions, Financial Aid, Registrat's, Graduate <u>http://www.mankato.mstus.edu/</u> -size 9K - 27 Nor 96	
Mississippi State University http://www.misstate.edu/ 66.35% Sit - 45920987 - 02003987 Northwestem University, NUInfo http://www.num.edu/ 06.15% Sit - 2202498 - 02005997 pext 10	St. Ambrose University Main Index: Academic Departments. Administrative Services. Campu News. Computing Services. Galvin Fine Arts Center. Internet Connections. Library <u>http://www.soul.od//</u> -size 2K - 4 Feb 27 University of Washington ECSEL Projects Image: Complexity of Washington ECSEL Projects	3 7
7/40		12

http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf

Unsupervised Graph-Based Ranking

- Instead of supervision, what if we have graph between examples?
 - Every paper is a node, and every citation is an edge.
 - Every Facebook user is a node, and every "friendship" is an edge.
 - Every webpage is a node, and every web-link is an edge.







https://en.wikipedia.org/wiki/Scale-free_network http://blog.revolutionanalytics.com/2010/12/facebooks-

http://mathematica.stackexchange.com/questions/11673/how-to-play-with-facebook-data-inside-mathematica

Unsupervised Graph-Based Ranking

- Instead of supervision, what if we have graph between examples?
 - Every paper is a node, and every citation is an edge.
 - Every Facebook user is a node, and every "friendship" is an edge.
 - Every webpage is a node, and every web-link is an edge.
- Key idea: use links (edges) to predict important of nodes.
- Many link analysis methods, usually with recursive definitions:
 A journal is "influential" if it is cited by "influential" journals.
- We will discuss PageRank, Google's original ranking algorithm.

PageRank

- Wikipedia's cartoon illustration of PageRank:
 - Large face => higher rank.

- Key ideas:
 - Important webpages are linked from other important webpages.
 - Link is more meaningful if a webpage has few links.



Random Walk View of PageRank

- PageRank algorithm can be interpreted as a random walk:
 - At time t=0, start at a random webpage.
 - At time t=1, follow a random link on the current page.
 - At time t=2, follow a random link on the current page.
- PageRank:
 - Probability of landing on page as t-> ∞ .
- Obvious problem:
 - Pages with no in-links have a rank of 0.
 - Algorithm can get "stuck" in part of the graph.



Random Walk View of PageRank

- Fix: add small probability of going to a random webpage at time 't'.
- Damped PageRank algorithm:
 - At time t=0, start at a random webpage.
 - At time t=1:
 - With probability α : go to a random webpage.
 - With probability (1- α): follow a random link on the current page.
 - At time t=2, follow a random link on the current page.
 - With probability α : go to a random webpage.
 - With probability (1- α): follow a random link on the current page.
- PageRank:
 - Probability of landing on page as t-> ∞ .

Markov Chains

- This random walk algorithm is a special case of a Markov chain:
 - Most common framework for modeling sequences.
 - Bioinformatics, physics/chemistry, speech recognition, predator-prey models, language tagging/generation, computing integrals, economic models, tracking missiles/players, modeling music.



Part 6: Markov Chains

- Random walk algorithm is a special case of a Markov chain.
- Markov chain ingredients:
 - State space:
 - Set of possible states we can be in at time 't' (webpages for PageRank).
 - Initial probabilities:
 - p(x₀ = s) that we start in state 's' at time 0.
 - Transition probabilities:
 - $p(x_t = s | x_{t-1} = s')$ that we move to state s to state s'.
- This model makes the Markov assumption:
 - Our state time at 't' only depends on the state at time t-1.
- Often assume homogeneous chain: transitions constant with 't'.

Markov Chains

- 3 things you can do with Markov chains:
 - You can simulate sequences:
 - Sample state x₀ from initial probabilities.
 - For t = 1:d
 - Sample x_t from transition probabilities.
 - Compute marginal probability of being in state 's' at time 't':
 - At time 0, just use the initial probabilities.
 - At time t > 0, marginalization and product rules gives recursive formula:

$$p(x_{t} = s) = \sum_{s'=1}^{|s|} p(x_{t} = s, x_{t-1} = s') = \sum_{s'=1}^{|s|} p(x_{t} = s | x_{t-1} = s') p(x_{t-1} = s') p(x_{t-1} = s')$$

$$matrix$$

$$matrix$$

$$multip kcaki$$

- Compute stationary distribution (PageRank):
 - P(xt = s) as 't' goes to infinity. ZExist, and is unique under certain assumptions works

PageRank Computation

- Monte Carlo method for computing PageRank:
 - Just run the random walk algorithm a really long time.
 - Count the number of times you visit each webpage.
 - Maybe include a "burn in" time at the start where you don't count pages.
 - Can parallelize by using random 'm' independent surfers.
 - Intuitive but slow.
- It can also be solved analytically with SVD:
 But O(n³) for 'n' webpages.
- Google's approach is the power method:
 - Repeated multiplication by transition matrix: O(nLinks) per iteration.

Application: Game of Thrones

- PageRank can be used for other applications.
- "Who is the main character in the Game of Thrones books?"



http://qz.com/650796/mathematicians-mapped-out-every-game-of-thrones-relationship-to-find-the-main-character

Ranking Discussion

- Modern ranking methods are more advanced:
 - Guarding against methods that exploit algorithm.
 - Removing offensive/illegal content.
 - Personalized recommendations.
 - Take into account that you often only care about top rankings.
 - Define losses that are not additive across ratings.
 - "Precision at k": if we return k documents, how many are relevant?
 - "Average precision": precision at k averaged across values of 'k'.
 - You can still define losses based on probability ratios:
 - But you get exponential number of terms, need more advanced optimization tricks.
 - Also work on diversity of rankings:
 - E.g., divide objects into sub-topics and do weighted 'covering' of topics.
 - Persistence/freshness as in recommender systems.

Summary

- Ranking orders objects based information about relationships.
- Supervised ranking contains explicit label information:
 - Item relevance assumes we have "scores" y_i.
 - Pairwise preferences assume we have relative rankings $y_i > y_j$.
- Probability ratios allow us to define more loss functions.
- Graph-based ranking uses links to solve ranking queries.
 PageRank is based on a model of a random web user.
- Markov chains are a general framework for modeling sequences.

• Next time: finding all the cat videos on YouTube.