RANKING

David Kauchak
CS 451 – Fall 2013

Admin

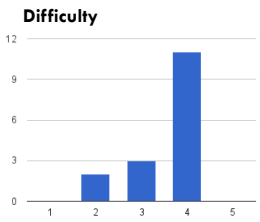
Assignment 4

Assignment 5

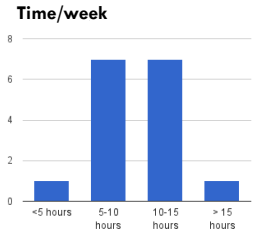
Course feedback

Thanks!

Course feedback



Course feedback



Course feedback

Experiments/follow-up questions

More interesting data sets

More advanced topics I'm *hoping* to get to:

- large margin classifiers
- probabilistic modeling
- Unsupervised learning/clustering
- Ensemble learning
- Collaborative filtering
- MapReduce/Hadoop

Java tip for the day

static

```
ArrayList<Example> data = dataset.getData();
```

How can I iterate over it?

Java tip for the day

```
ArrayList<Example> data = dataset.getData();

for( int i = 0; i < data.size(); i++ ){
        Example ex = data.get(i)
}
```

OR

```
// can do on anything that implements the Iterable interface
for( Example ex: data ){

}
```

An aside: text classification

Raw data | labels


Chardonnay


Pinot Grigio


Zinfandel

Text: raw data

Raw data | labels | Features?


Chardonnay


Pinot Grigio


Zinfandel

Feature examples

Raw data | labels | Features


Chardonnay

Clinton said pinot repeatedly last week on tv, "pinot, pinot, pinot"


Pinot Grigio

(1, 1, 1, 0, 0, 1, 0, 0, ...)

pinot clinton said california across tv wrong capital


Zinfandel

Occurrence of words

Feature examples

Raw data | labels | Features


Chardonnay

Clinton said pinot repeatedly last week on tv, "pinot, pinot, pinot"


Pinot Grigio

(4, 1, 1, 0, 0, 1, 0, 0, ...)

pinot clinton said california across tv wrong capital

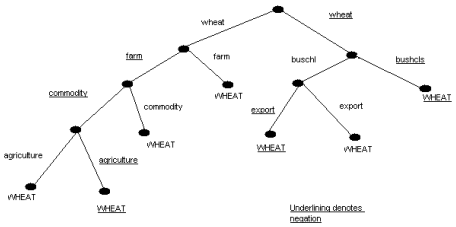

Zinfandel

Frequency of word occurrences

This is the representation we're using for assignment 5

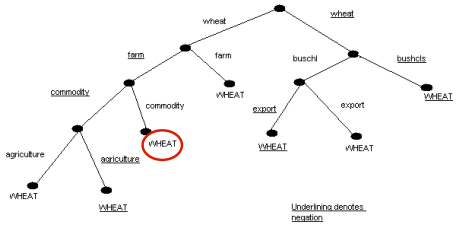
Decision trees for text

Each internal node represents whether or not the text has a particular word



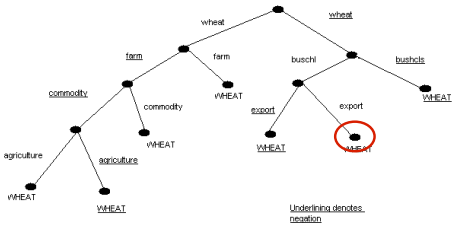
Decision trees for text

wheat is a commodity that can be found in states across the nation

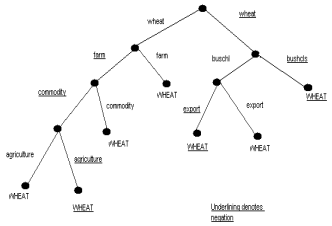


Decision trees for text

The US views technology as a commodity that it can export by the buschl.



Printing out decision trees




```
(wheat
  (buschl
    predict: not wheat
    (export
      predict: not wheat
      predict: wheat))
  (farm
    (commodity
      (agriculture
        predict: not wheat
        predict: wheat)
      predict: wheat)
    predict: wheat))
```

Ranking problems

Suggest a simpler word for the word below:

vital

Suggest a simpler word

Suggest a simpler word for the word below:

vital

word	frequency
important	13
necessary	12
essential	11
needed	8
critical	3
crucial	2
mandatory	1
required	1
vital	1

Suggest a simpler word

Suggest a simpler word for the word below:

acquired

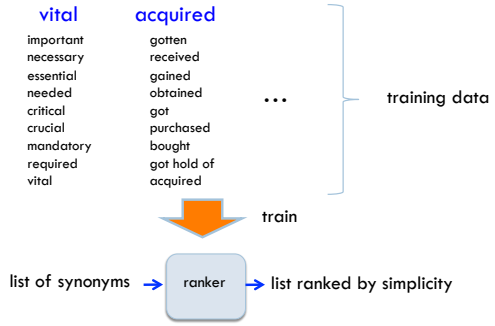
Suggest a simpler word

Suggest a simpler word for the word below:

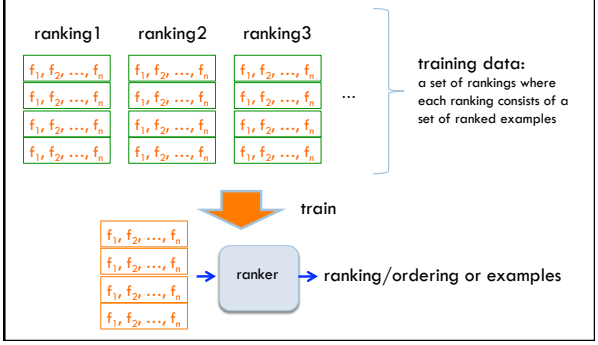
acquired

word	frequency
gotten	12
received	9
gained	8
obtained	5
got	3
purchased	2
bought	2
got hold of	1
acquired	1

Suggest a simpler word



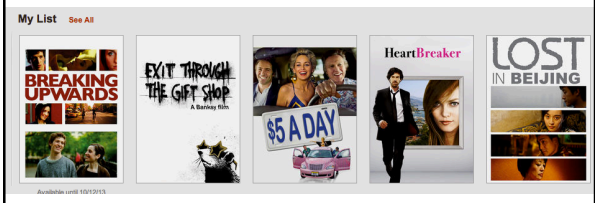
Ranking problems in general



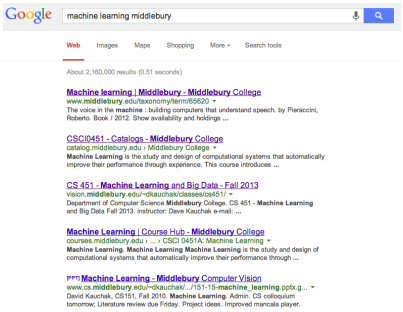
Ranking problems in general

Real-world ranking problems?

Netflix My List



Search



Ranking Applications

reranking N-best output lists

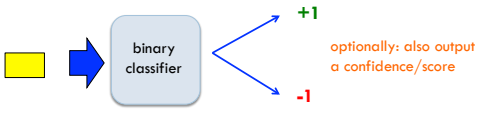
- machine translation
- computational biology
- parsing
- …

flight search

…

Black box approach to ranking

Abstraction: we have a generic binary classifier, how can we use it to solve our new problem



Can we solve our ranking problem with this?

Predict better vs. worse

Train a classifier to decide if the first input is better than second:
- Consider all possible pairings of the examples in a ranking
- Label as positive if the first example is higher ranked, negative otherwise

ranking 1

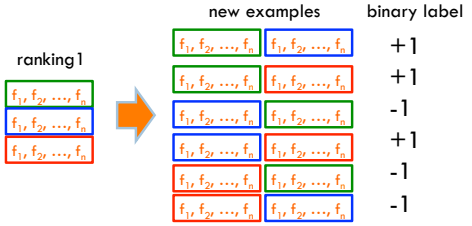
$f_1, f_2, \ldots, f_n$

$f_1, f_2, \ldots, f_n$

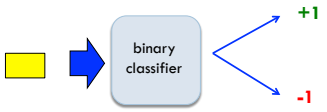
$f_1, f_2, \ldots, f_n$

Predict better vs. worse

Train a classifier to decide if the first input is better than second:
- Consider all possible pairings of the examples in a ranking
- Label as positive if the first example is higher ranked, negative otherwise

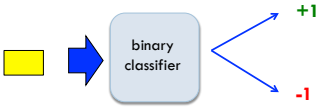


Predict better vs. worse

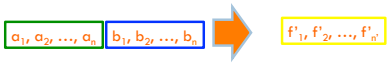


Our binary classifier only takes one example as input

Predict better vs. worse



Our binary classifier only takes one example as input



How can we do this?
We want features that compare the two examples.

Combined feature vector

Many approaches! Will depend on domain and classifier

Two common approaches:

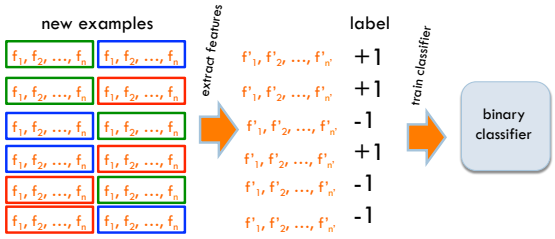
1. difference:

$$f'_i = a_i - b_i$$

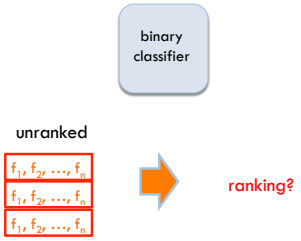
2. greater than/less than:

$$f'_i = \begin{cases} 1 & if\ a_i > b_i \\ 0 & otherwise \end{cases}$$

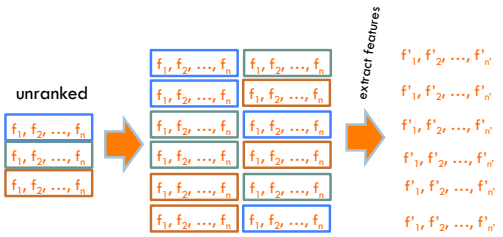
Training



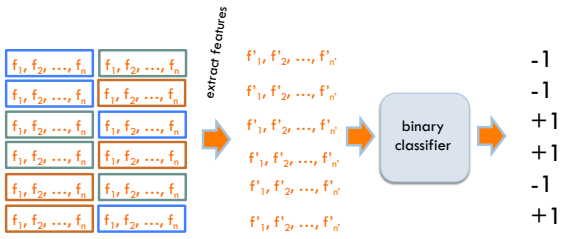
Testing



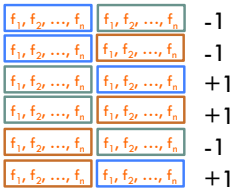
Testing



Testing



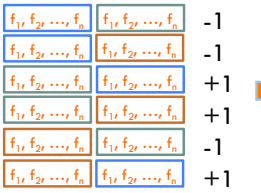
Testing



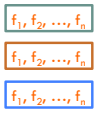
What is the ranking?
Algorithm?

Testing

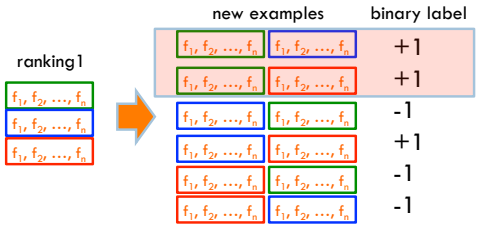
for each binary example e_{jk}:
label[j] += $f_{jk}(e_{jk})$
label[k] -= $f_{jk}(e_{jk})$



rank according to label scores

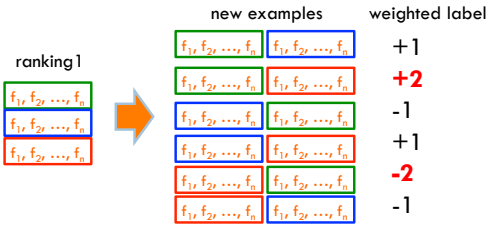


An improvement?



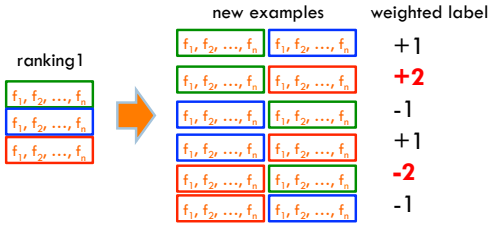
Are these two examples the same?

Weighted binary classification



Weight based on ***distance*** in ranking

Weighted binary classification



In general can weight with any consistent distance metric

Can we solve this problem?

Testing

If the classifier outputs a confidence, then we've learned a *distance* measure between examples

During testing we want to rank the examples based on the learned distance measure

Ideas?

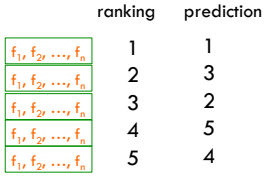
Testing

If the classifier outputs a confidence, then we've learned a *distance* measure between examples

During testing we want to rank the examples based on the learned distance measure

Sort the examples and use the output of the binary classifier as the similarity between examples!

Ranking evaluation



Ideas?

Idea 1: accuracy

	ranking	prediction
$f_1, f_2, \ldots, f_n$	1	1
$f_1, f_2, \ldots, f_n$	2	3
$f_1, f_2, \ldots, f_n$	3	2
$f_1, f_2, \ldots, f_n$	4	5
$f_1, f_2, \ldots, f_n$	5	4

1/5 = 0.2

Any problems with this?

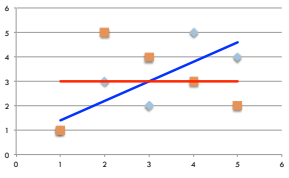
Doesn't capture "near" correct

	ranking	prediction	prediction
$f_1, f_2, \ldots, f_n$	1	1	1
$f_1, f_2, \ldots, f_n$	2	3	5
$f_1, f_2, \ldots, f_n$	3	2	4
$f_1, f_2, \ldots, f_n$	4	5	3
$f_1, f_2, \ldots, f_n$	5	4	2

1/5 = 0.2

Idea 2: correlation

ranking	prediction	prediction
1	1	1
2	3	5
3	2	4
4	5	3
5	4	2



Look at the correlation between the ranking and the prediction