Latent Semantic Indexing

Introduction to
Artificial Intelligence

COS302

Michael L. Littman

Fall 2001

Administration

Example analogies…

And-or Proof

out(<u>x</u>) = g($\text{sum}_k\ w_k\ x_k$)

w_1=10, w_2=10, w_3=-10 $x_1 + x_2 + \sim x_3$

Sum for 110?

Sum for 001?

Generally? <u>b</u>=110, 20 -10 $\text{sum}_i\ |b_i - x_i|$

What happens if we set

- w_0=10?
- w_0 =-15?

LSI Background Reading

Landauer, Laham, Foltz (1998). Learning human-like knowledge by Singular Value Decomposition: A Progress Report. *Advances in Neural Information Processing Systems* 10, (pp. 44-51)

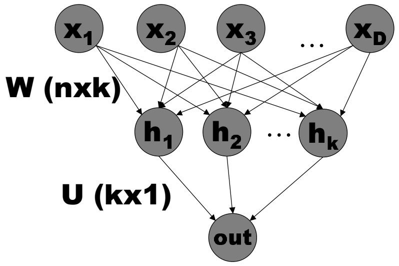
http://lsa.colorado.edu/papers/nips.ps

Outline

Linear nets, autoassociation

LSI: Cross between IR and NNs

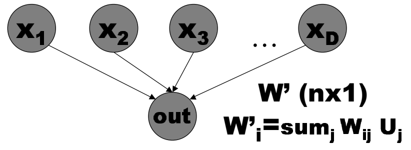
Purely Linear Network



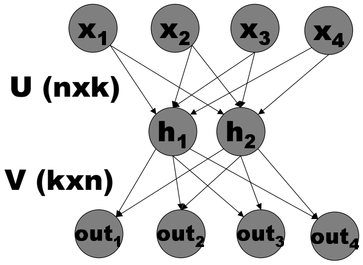
What Does It Do?

$\text{out}(\underline{x}) = \text{sum}_j\ (\text{sum}_i\ x_i\ W_{ij})\ U_j$

$= \text{sum}_i\ x_i\ (\text{sum}_j\ W_{ij}\ U_j\ )$



Can Other Layers Help?



Autoassociator

x_1	x_2	x_3	x_4	h_1	h_2	y_1	y_2	y_3	y_4
1	0	0	0			1	0	0	0
0	1	0	0			0	1	0	0
0	0	1	0			0	0	1	0
0	0	0	1			0	0	0	1

Applications

Autoassociators have been used for data compression, feature discovery, and many other tasks.

U matrix encodes the inputs into k features

How train?

SVD

Singular value decomposition provides another method, from linear algebra.

Training data M is nxm (input features by examples)

$$M = U\ \Sigma^{2}{}_{k}\ V^{T}$$

$U^TU = I$, $V^TV = I$, Σ diagonal

Dimension Reduction

Finds least squares best U (nxk, free k)

Rows of U map input features to encoded features (instance is sum)

Closely related to

- symm. eigenvalue decomposition,
- factor analysis
- principle component analysis

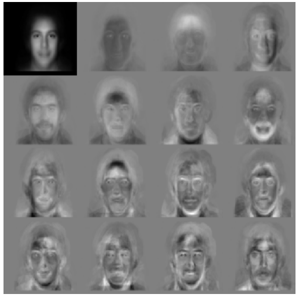
Subroutine in many math packages.

SVD Applications

Eigenfaces

Handwriting recognition

Text applications…



LSI/LSA

Latent semantic indexing is the application of SVD to IR.

Latent semantic analysis is the more general term.

Features are words, examples are text passages.

Latent: Not visible on the surface

Semantic: Word meanings

Running LSI

Learns new word representations!

Trained on:

- 20,000-60,000 words
- 1,000-70,000 passages

Use k=100-350 hidden units

Similarity between vectors computed as cosine.

Step by Step

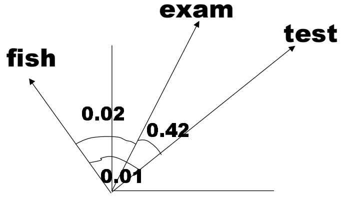
1. M_{ij} rows are words, columns are passages: filled w/ counts
2. Transformation of matrix:

$$\frac{\log(M_{ij}+1)}{-\text{sum}_j\,((M_{ij}/\text{sum}_j M_{ij})\log(M_{ij}/\text{sum}_j M_{ij})}$$

3. SVD computed: $M=U\Sigma V^T$
4. Best k components of rows of U kept as word representations.

Geometric View

Words embedded in high-d space.



Comparison to VSM

A:The feline climbed upon the roof

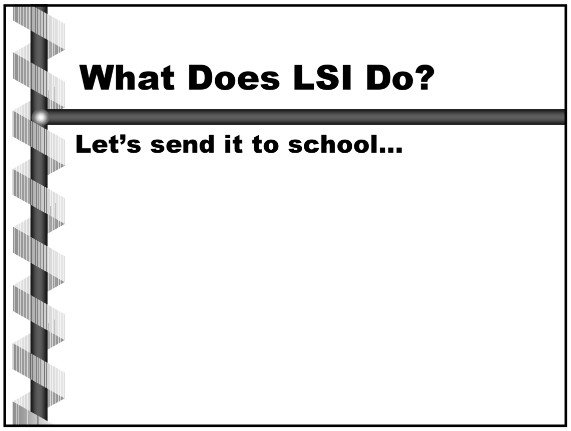
B:A cat leapt onto a house

C:The final will be on a Thursday

How similar?

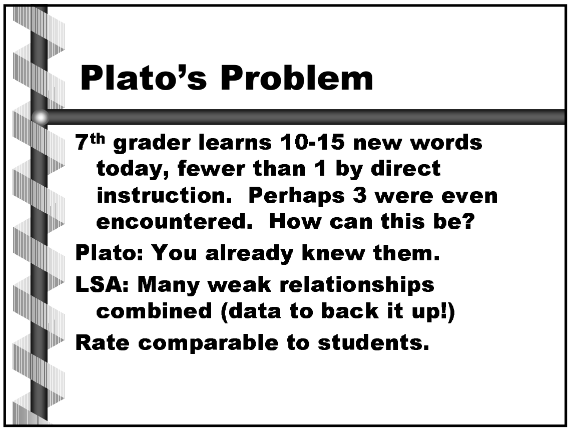
- Vector space model: sim(A,B)=0
- LSI: sim(A,B)=.49>sim(A,C)=.45

Non-zero sim with no words in common by overlap in reduced representation.



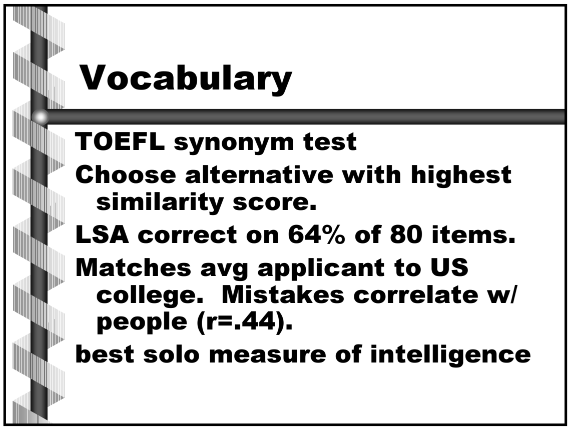
What Does LSI Do?

Let's send it to school…



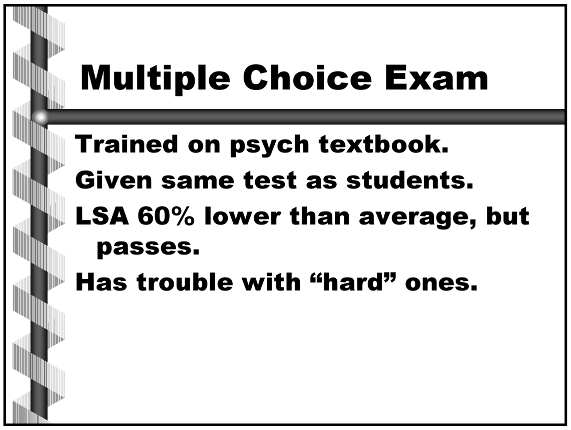
Plato's Problem

- 7th grader learns 10-15 new words today, fewer than 1 by direct instruction. Perhaps 3 were even encountered. How can this be?
- Plato: You already knew them.
- LSA: Many weak relationships combined (data to back it up!)
- Rate comparable to students.



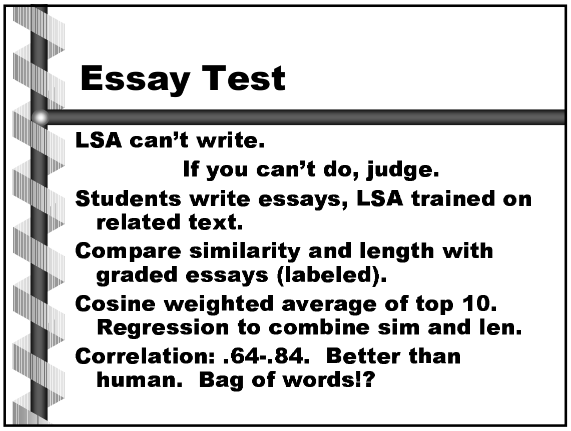
Vocabulary

- TOEFL synonym test
- Choose alternative with highest similarity score.
- LSA correct on 64% of 80 items.
- Matches avg applicant to US college. Mistakes correlate w/ people (r=.44).
- best solo measure of intelligence



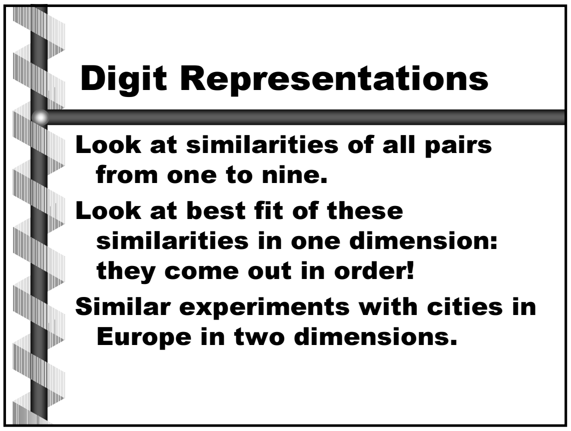
Multiple Choice Exam

- Trained on psych textbook.
- Given same test as students.
- LSA 60% lower than average, but passes.
- Has trouble with "hard" ones.



Essay Test

- LSA can't write.
  - If you can't do, judge.
- Students write essays, LSA trained on related text.
- Compare similarity and length with graded essays (labeled).
- Cosine weighted average of top 10. Regression to combine sim and len.
- Correlation: .64-.84. Better than human. Bag of words!?



Digit Representations

- Look at similarities of all pairs from one to nine.
- Look at best fit of these similarities in one dimension: they come out in order!
- Similar experiments with cities in Europe in two dimensions.

Word Sense

The chemistry student knew this was not a good time to forget how to calculate volume and mass.

heavy? .21

church? .14

LSI picks best p<.001

More Tests

- Antonyms just as similar as syns. (Cluster analysis separates.)
- LSA correlates .50 with children and .32 with adults on word sorting (misses grammatical classification).
- Priming, conjunction error: similarity correlates with strength of effect

Conjunction Error

Linda is a young woman who is single, outspoken…deeply concerned with issues of discrimination and social justice

Is Linda a feministic bank teller?

Is Linda a bank teller?

80% rank former has higher. Can't be!

Pr(f bt | Linda) = Pr(bt | Linda) Pr(f | Linda, bt)

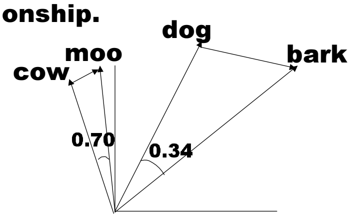
LSApplications

1. Improve IR.
2. Cross-language IR. Train on parallel collection.
3. Measure text coherency.
4. Use essays to pick educational text.
5. Grade essays.

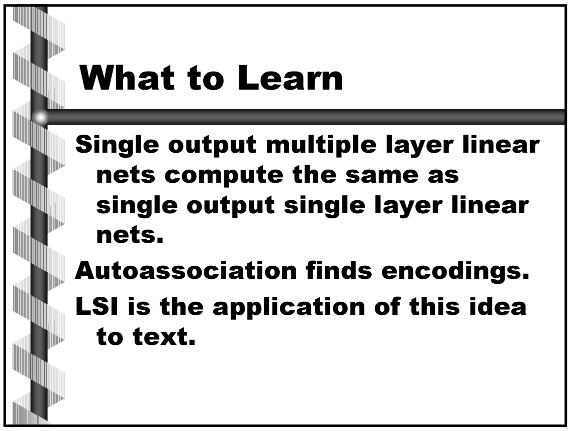
Demos at http://LSA.colorado.edu

Analogies

Compare difference vectors: geometric instantiation of relationship.



LSA Motto? (AT&T Cafeteria)

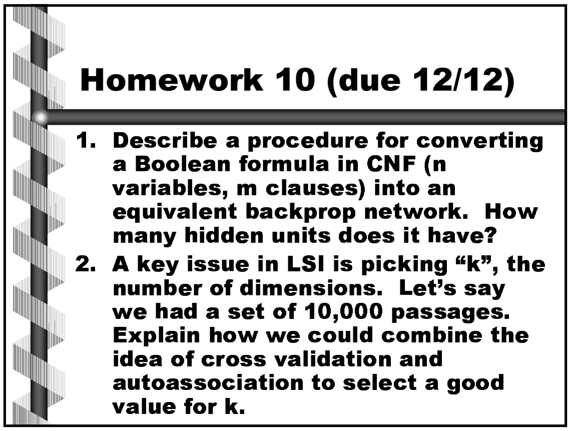


What to Learn

Single output multiple layer linear nets compute the same as single output single layer linear nets.

Autoassociation finds encodings.

LSI is the application of this idea to text.



Homework 10 (due 12/12)

1. Describe a procedure for converting a Boolean formula in CNF (n variables, m clauses) into an equivalent backprop network. How many hidden units does it have?
2. A key issue in LSI is picking “k”, the number of dimensions. Let’s say we had a set of 10,000 passages. Explain how we could combine the idea of cross validation and autoassociation to select a good value for k.