

SMAI-M20-L11: Semantic Representation for all: "Words", "Documents" "Consumers" and "Products"

C. V. Jawahar

IIIT Hyderabad

September 2, 2020



Class Review Questions

- Consider a vocabulary of size d . One hot representation of a word i is “1” at the location (index) corresponding to that word and zero elsewhere.

Given a document that contains P words, $\mathbf{w}_1, \dots, \mathbf{w}_P$, we compute

$$\mathbf{x} = \sum_{i=1}^P \mathbf{w}_i$$

- Consider a document is represented by a histogram of the words in the document. \mathbf{h} i.e., h_i is the number of occurrence of the i th word in the document.

If the document is paraphrased in certain manner, does the representation change ? (invariant = does not vary/change)

- Can we address paraphrasing if we had pre-multiplied by a certain similarity matrix.

$$\sum (\omega^T x_n) (x_n^T c) \quad \propto \quad \sum \omega^T \left(\sum_n x_n x_n^T \right) c$$

$$[\quad] [\quad]$$

Recap:

- Problem Space:
 - Learn a function $y = f(\mathbf{W}, \mathbf{x})$ from the data.
 - Learn useful features
- Supervised Learning:
 - Notion of Training, Validation and Testing
 - Loss Function and Optimization
 - Need of Generalization and Worry of Overfitting
 - Occam's razor and role of model complexity
 - Balancing between Bias and Variance
 - Estimating error using validation set.
- Classification Algorithms:
 - Nearest Neighbour Algorithm
 - Linear Classification; Linear Regression
 - Decide as ω_1 if $P(\omega_1|\mathbf{x}) \geq P(\omega_2|\mathbf{x})$ else ω_2
 - Performance Metrics
- Mathematical Foundations: Linear Algebra, Probability, Optimization
 - SVD, Eigen Decomposition, MLE
 - LSI,

Representation: Bag of Words and One-Hot

The Bag of Words Representation

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!



it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
...	...

h

Σ

$[000\underline{1}0000]^d$

Representation: Bag of Words and One-Hot

2007-01-23: State of the Union Address

George W. Bush (2001-)

abandon accountable affordable afghanistan africa aided ally anbar armed army baghdad bless challenges chamber chaos
choices civilians coalition commanders commitment confident confront congressman constitution corps debates deduction
deficit deliver democratic deploy dikembe diplomacy disruptions earmarks economy einstein elections eliminates
expand extremists falling faithful families freedom fuel funding god haven ideology immigration impose
insurgents iran **iraq** islam julie lebanon love madam marine math medicare moderation neighborhoods nuclear offensive
palestinian payroll province pursuing **qaeda** radical regimes resolve retreat rieman sacrifices science sectarian senate
september shia stays strength students succeed sunni tax territories **terrorists** threats uphold victory
violence violent war washington weapons wesley

Representation: Bag of Words and One-Hot

2007-01-23: State of the Union Address

George W. Bush (2001-)

abandon a
choices civ

deficit de
expand E

insurgent
palestinian

septembe
violence

1962-10-22: Soviet Missiles in Cuba

John F. Kennedy (1961-63)

abandon achieving adversaries aggression agricultural appropriate armaments **arms** assessments atlantic ballistic berlin
buildup burdens cargo college commitment communist constitution consumers cooperation crisis **cuba** dangers
declined **defensive** deficit depended disarmament divisions domination doubled **economic** education
elimination emergence endangered equals europe expand exports fact false family forum **freedom** fulfill gromyko
halt hazards **hemisphere** hospitals ideals independent industries inflation labor latin limiting minister **missiles**
modernization neglect **nuclear** oas obligation observer **offensive** peril pledged predicted purchasing quarantine quote
recession rejection republics retaliatory safeguard sites solution **soviet** space spur stability standby **strength**
surveillance tax territory treaty undertakings unemployment **war** warhead **weapons** welfare western widen withdraw

Representation: Bag of Words and One-Hot

2007-01-23: State of the Union Address

George W. Bush (2001-)

abandon a
choices ch
deficit de
expand e

1962-10-22: Soviet Missiles in Cuba

John F. Kennedy (1961-63)

abandon achieve advance aggression agricultural appropriate armaments arms assessments atlantic ballistic berlin

Insurgent
palestinian
septembe
violence

1941-12-08: Request for a Declaration of War

Franklin D. Roosevelt (1933-45)

buildup
declined d
elimination
halt hazar
modernizat

recession r
surveillan

abandoning acknowledge aggression aggressors airplanes armaments armed army assault assembly authorizations bombing
britain british cheerfully claiming constitution curtail december defeats defending delays democratic dictators disclose
economic empire endanger facts false forgotten fortunes france freedom fulfilled fullness fundamental gangsters
german germany god guam harbor hawaii hemisphere hint hitler hostilities immune improving indies innumerable
invasion islands isolate japanese labor metals midst midway navy nazis obligation offensive
officially pacific partisanship patriotism pearl peril perpetrated perpetual philippine preservation privilege reject
repaired resisting retain revealing rumors seas soldiers speaks speedy stamina strength sunday sunk supremacy tanks taxes
treachery true tyranny undertaken victory war wartime washington

This Lecture:

Micro-Lecture Videos

① Rank and Recommendation Systems

- Appreciate why it is low-rank and also incomplete.
- How do we compare two different “users” or two different “products”?
- How do we formulate the matrix completion problem?

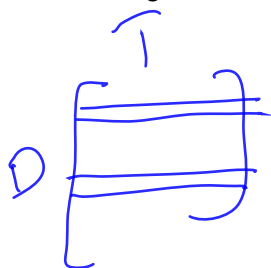
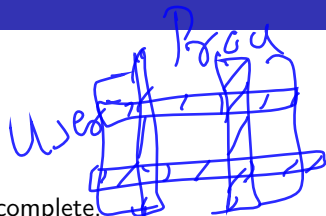
② Regularization in Regression

- Ridge and Lasso. (L2 and L1)
- Appreciate connection between regularization and overfitting.

③ Orthogonal line fitting

- Another Eigen vector problem
- Lead to PCA (to connect to the next lecture).

Questions? Comments?



$$\sum (\omega^T x_n - \underbrace{\omega^T \mu})^2$$

$$(-x_n - 1) (\omega^T x_n)^2$$

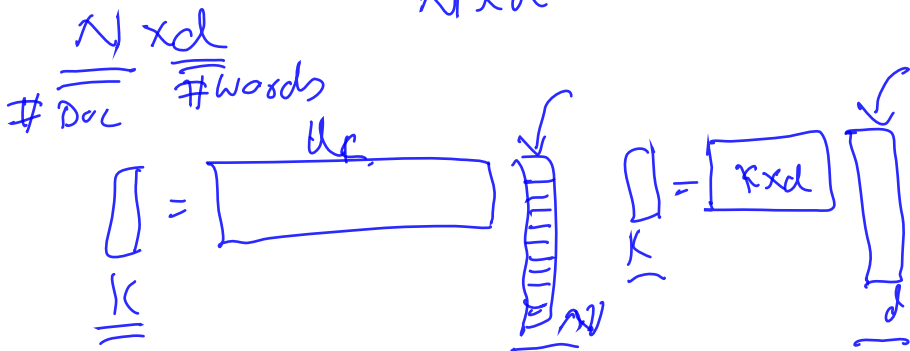
$$\sum \frac{\omega^T (x_n - \mu)}{a^T b} \frac{(x_n - \mu)^T}{b^T a}$$

$$\omega^T \left(\sum_n \underbrace{(x_n - \mu)} \underbrace{(x_n - \mu)^T} \right) \omega$$

$$\underline{\underline{\omega^T \Sigma \omega}}$$

$$\omega^T (x_0, x_1^T)$$

Blank

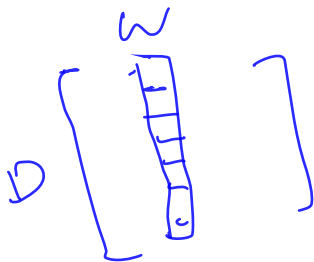


Discussion Point - I

$[0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]$

How do we represent words?

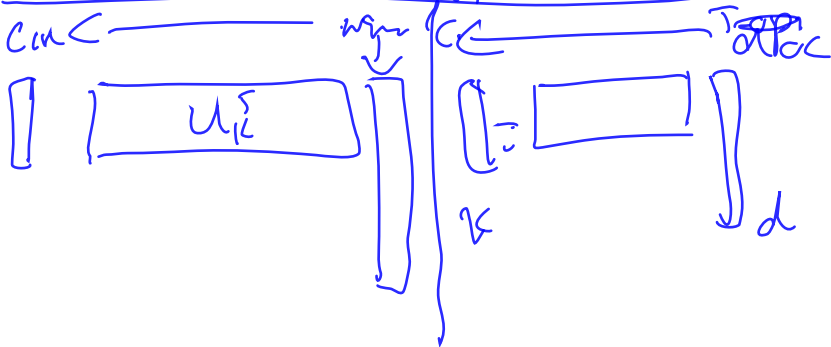
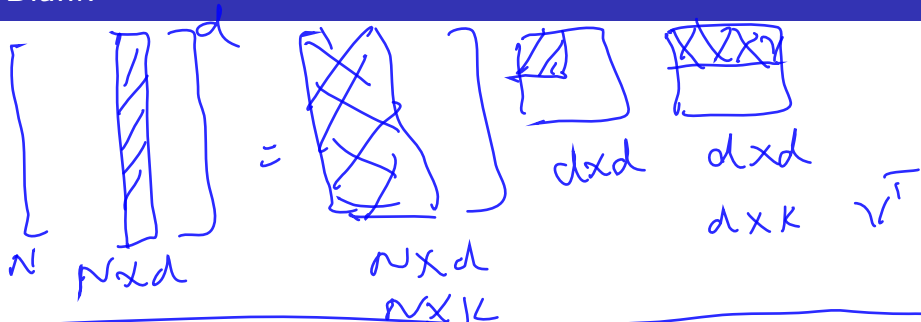
- 1 As one-hot vector
- 2 As columns of TD Matrix of size $N \times d$
- 3 After multiplying with U_k^T
- 4 Word2Vec



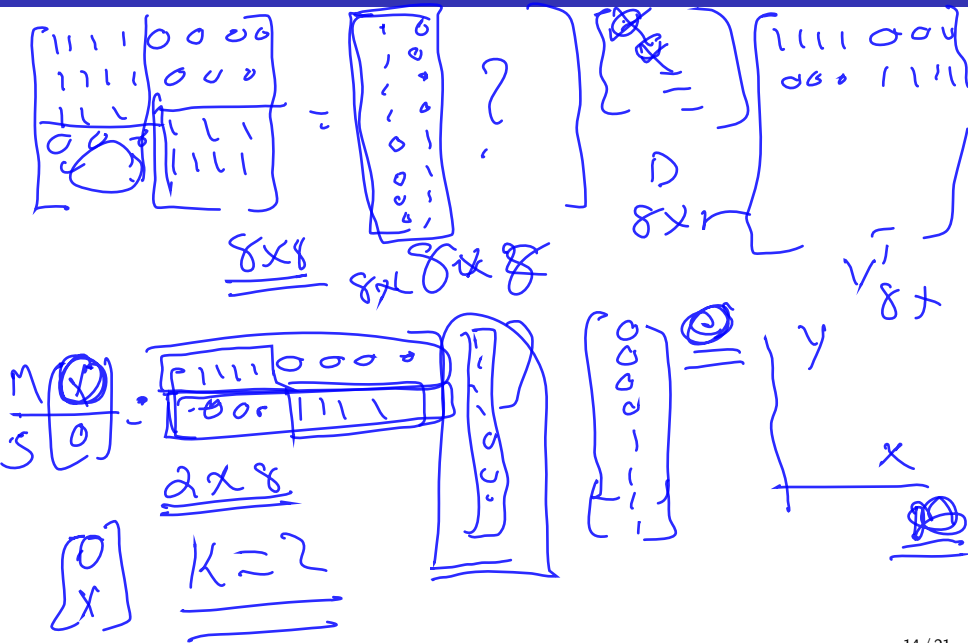
$$\begin{bmatrix} 0.5 \\ 0.1 \end{bmatrix} \sim \begin{bmatrix} 0.8 \\ 0.1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Below the second vector, there is a label 'x' and a vector $\begin{bmatrix} 0.1 \\ 0.8 \end{bmatrix}$ with an arrow pointing to the second vector.

Blank



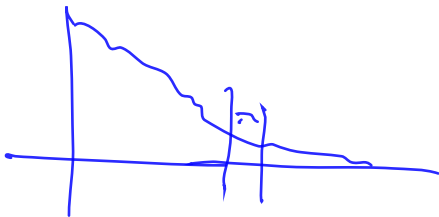
Blank



$K?$



K

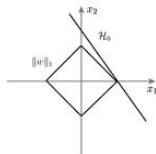


Consider the problem of finding solution to an ill-posed problem: Find the solution to $w_1x_1 + w_2x_2 = 10$ There are many and we prefer one that also minimize an L_p norm.

Regularization with L_p norm is popular. Why does L_1 norm induce sparsity?

Some Plots(from Internet)

A L1 regularization



B L2 regularization

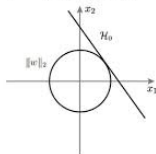
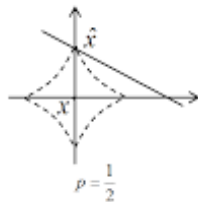
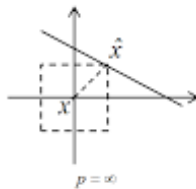
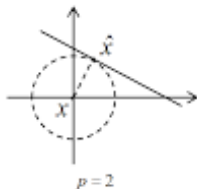
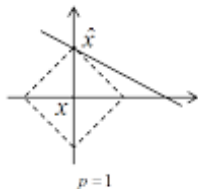
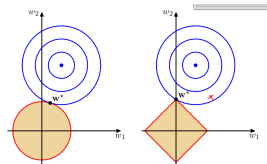


Figure 3.4 Plot of the contours of the unregularized error function (blue) along with the constraint region (3.30) for the quadratic regularizer $q = 2$ on the left and the lasso regularizer $q = 1$ on the right, in which the optimum value for the parameter vector w is denoted by w^* . The lasso gives a **sparse** solution in which $w_1^* = 0$.



$$p = \infty$$



$$p = 2$$



$$p = 1$$



$$0 < p < 1$$



$$p = 0$$

What Next:?

- ① PCA and Dimensionality Reduction
- ② Bayesian Optimal (Cont.)
- ③ What are good features?