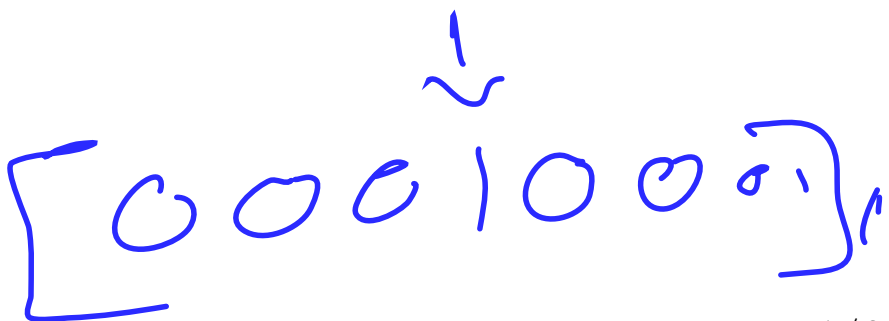
SMAI-M20-L11: Semantic Representation for all: "Words", "Documents" "Consumers" and "Products"

C. V. Jawahar

IIIT Hyderabad

September 2, 2020



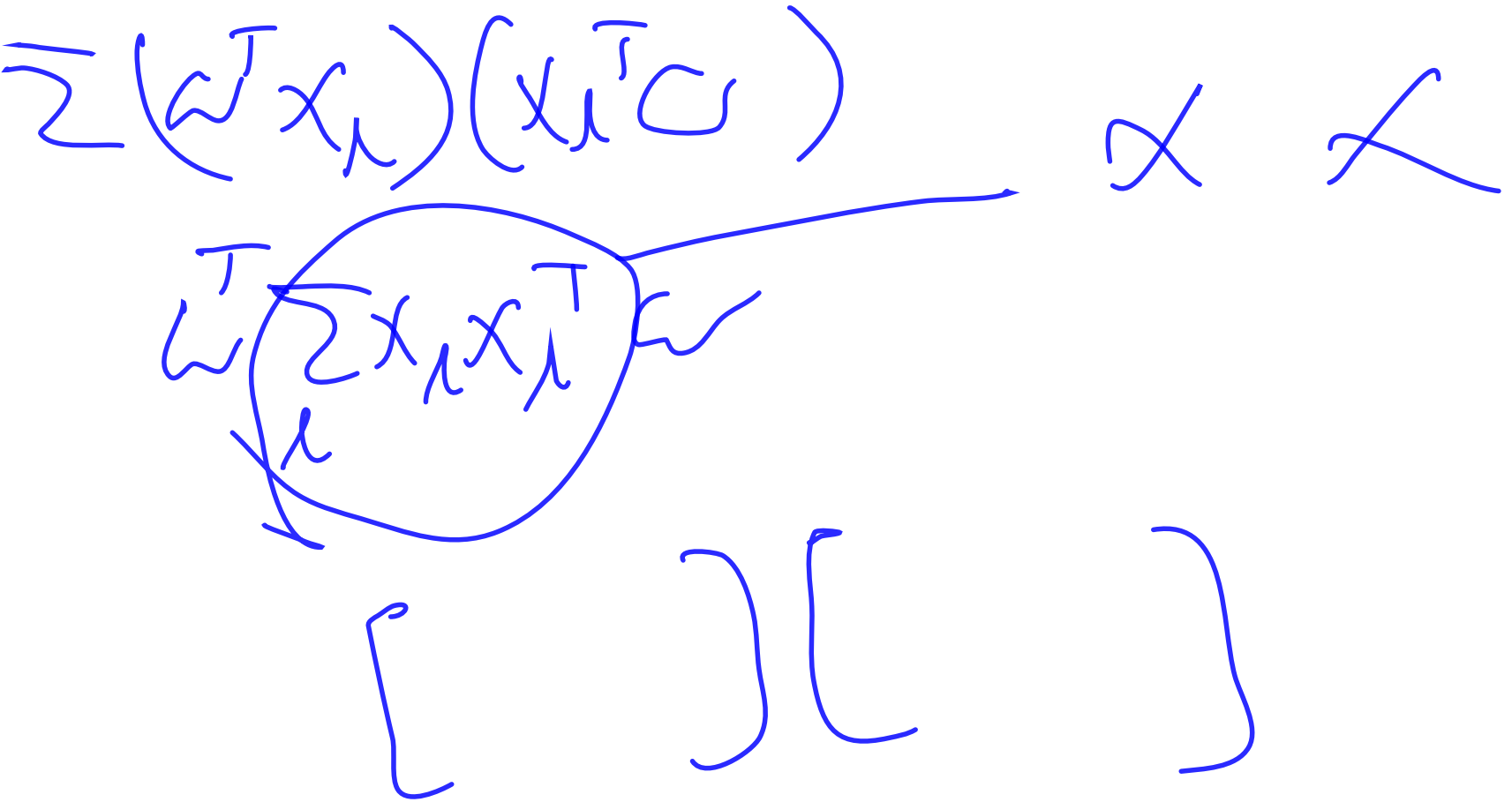
Class Review Questions

- 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 else where.
  Given a document that contains P words, $\mathbf{w_1}, \ldots, \mathbf{w_P}$, 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.

Blank



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
…	…

Doc

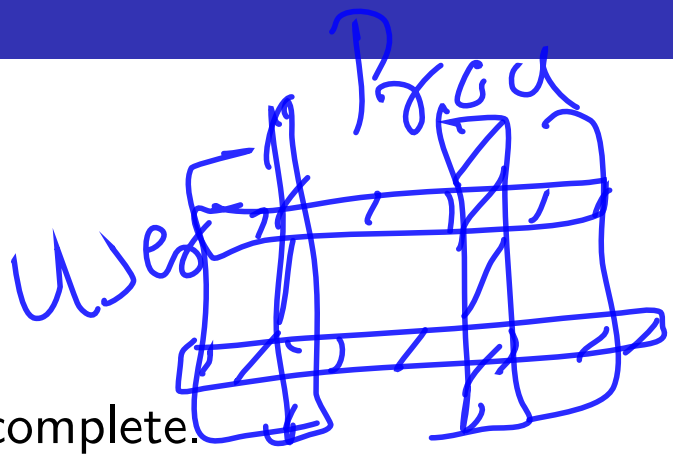
h Σ word [0 0 0 1 0 0 0 0 ...] d

Representation: Bag of Words and One-Hot

Representation: Bag of Words and One-Hot

Representation: Bag of Words and One-Hot

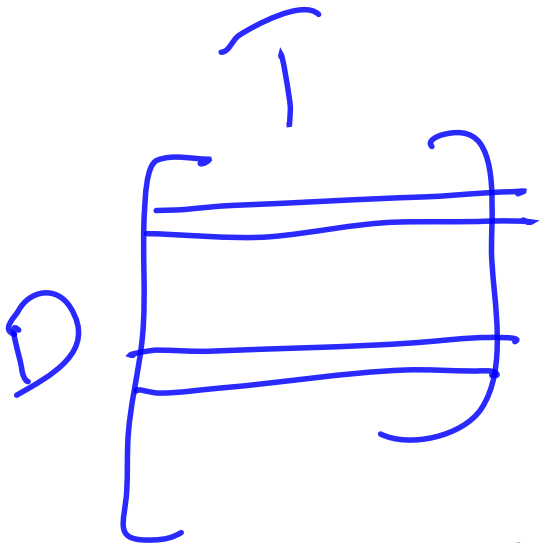
This Lecture:



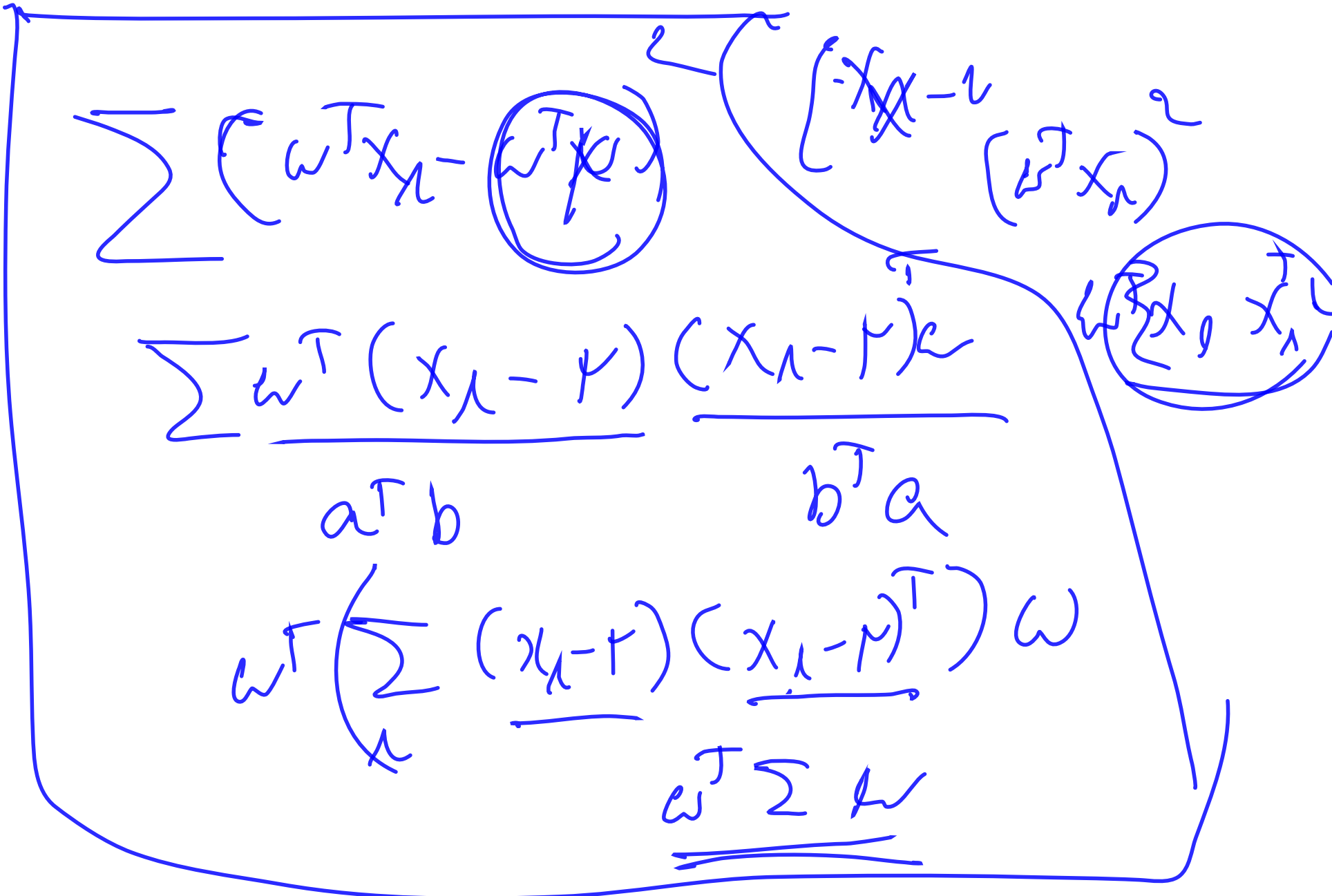
Micro-Lecture Videos

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).

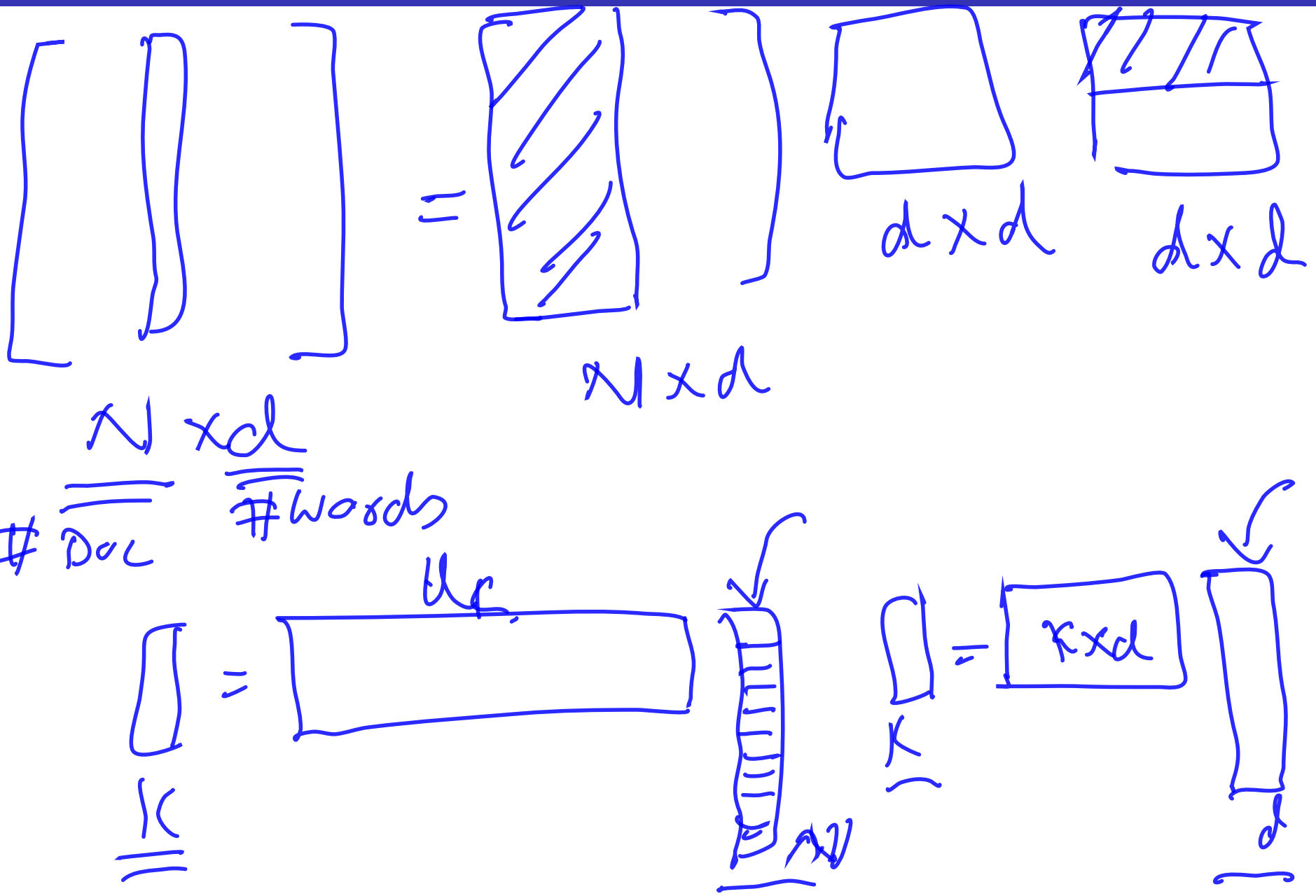
Questions? Comments?



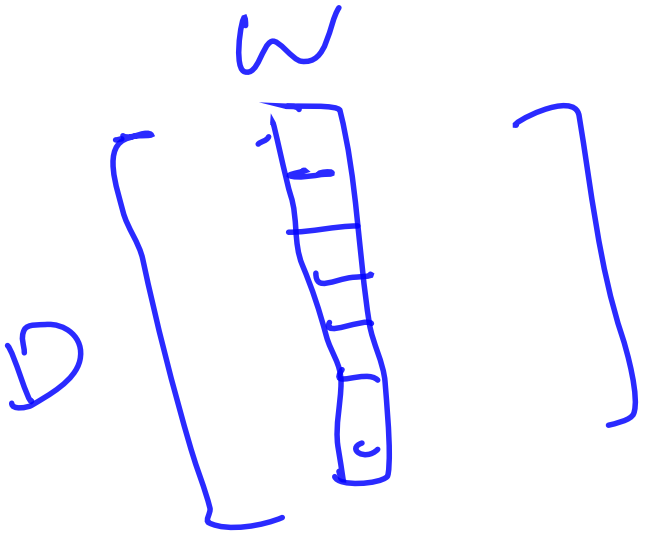
Blank



Blank

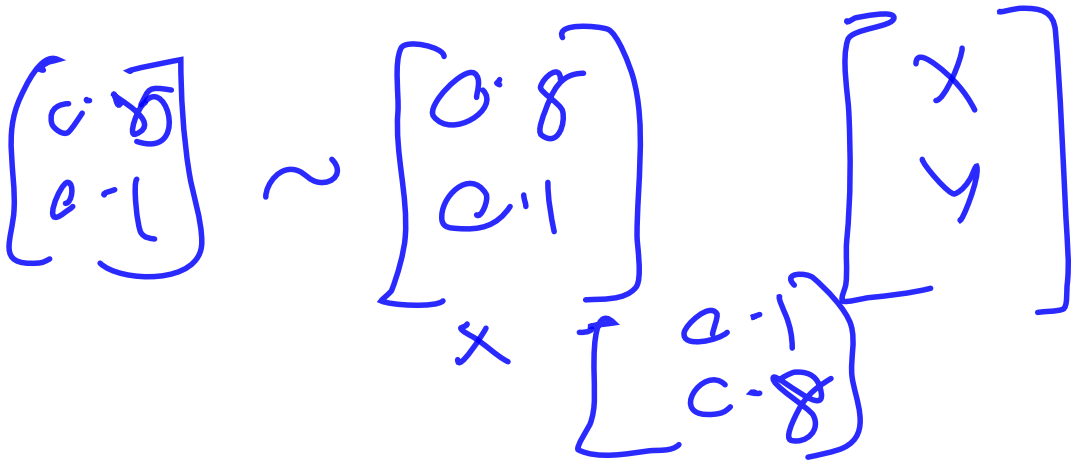


Discussion Point - I

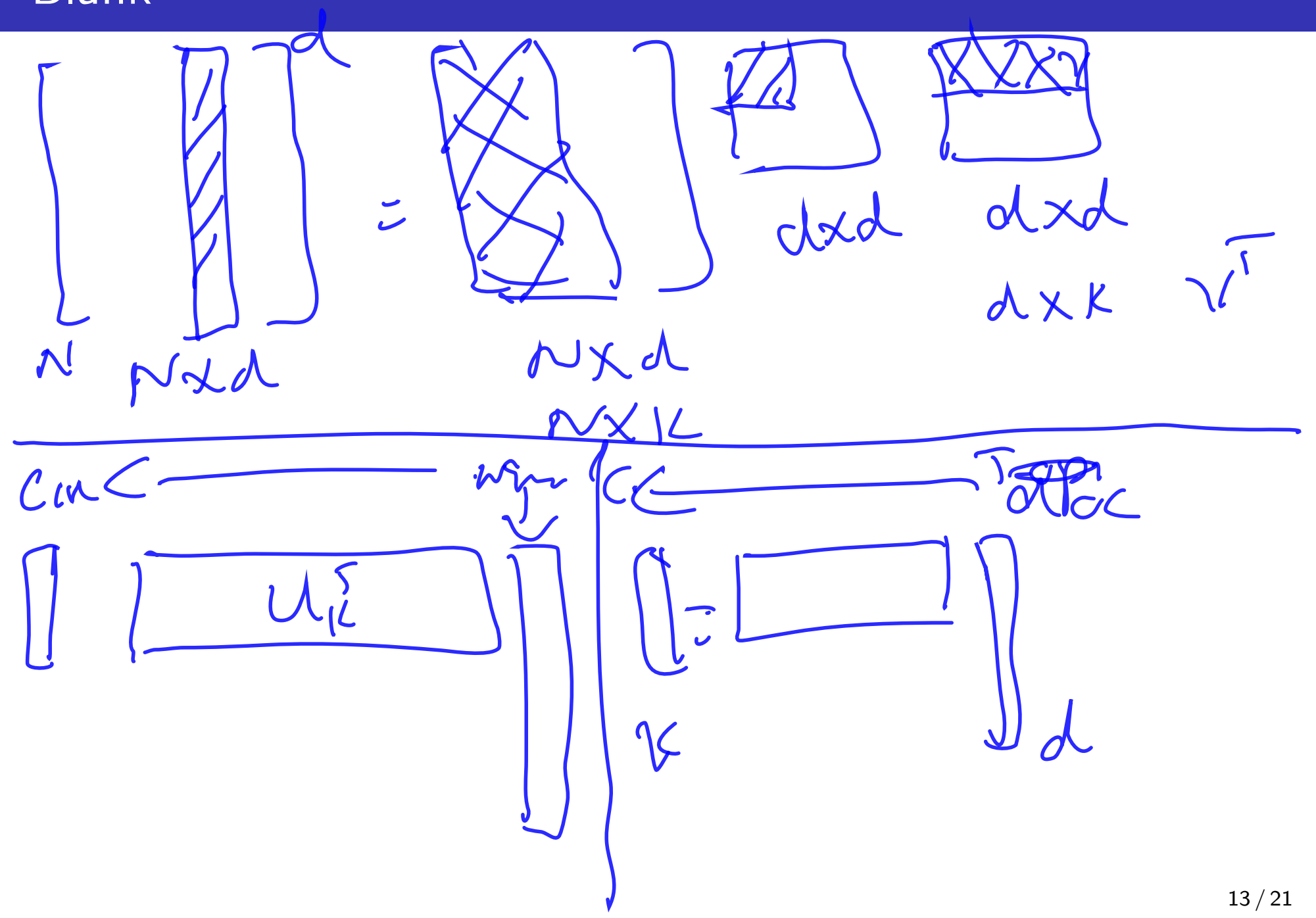


How do we represent words?

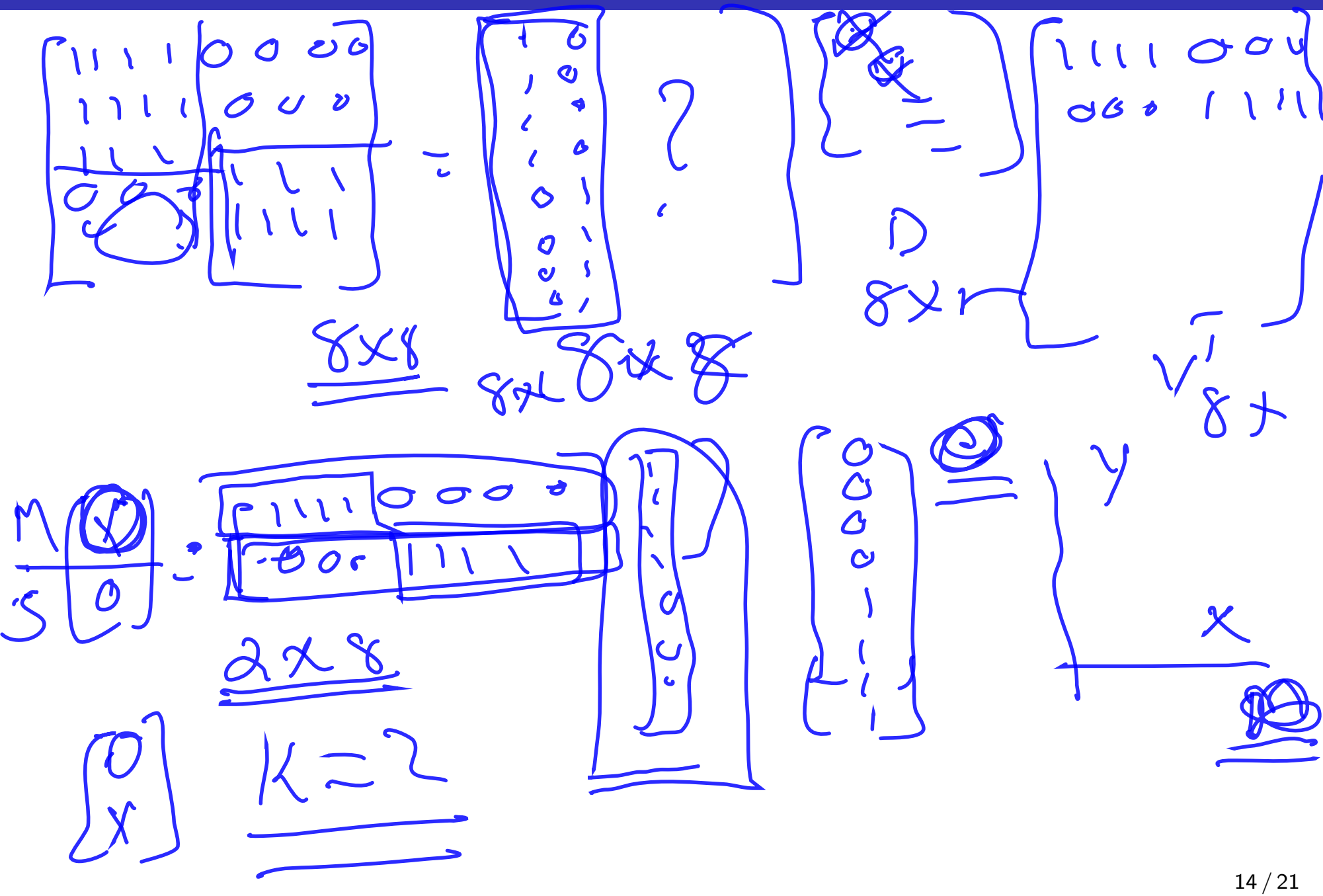
1. As one-hot vector
2. As columns of TD Matrix of size $N \times d$
3. After multiplying with U_k^T
4. Word2Vec



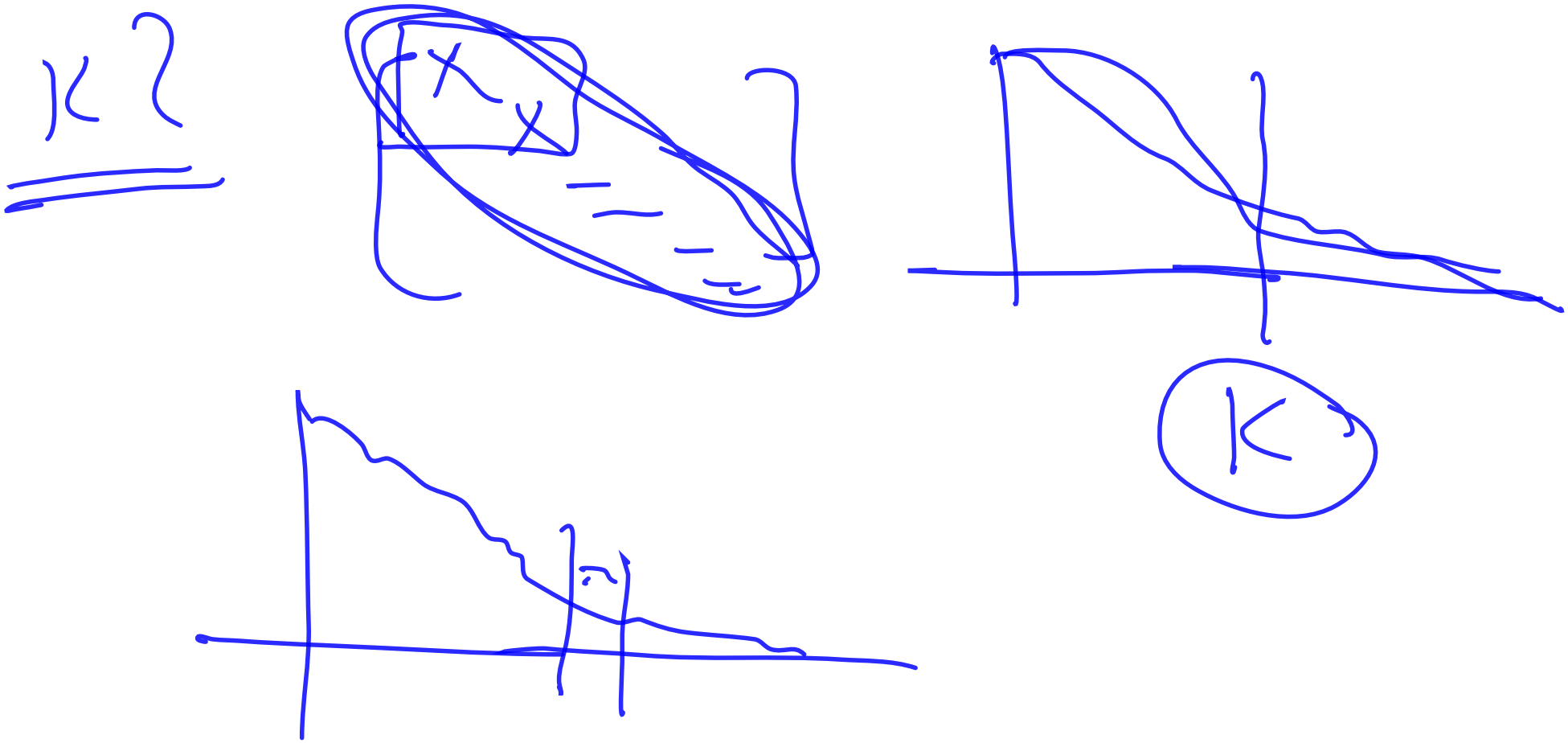
Blank



Blank



Blank



Discussions Point -II

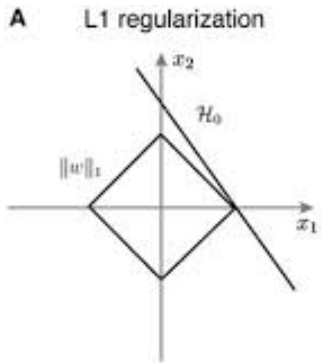
Consider the problem of finding solution to an ill-posed problem: Find the solution to $w_1 x_1 + w_2 x_2 = 10$ There are many and we prefer one that also minimize an Lp norm.

Regularization with Lp norm is popular. Why does L1 norm induce sparsity?

Blank

Blank

Some Plots(from Internet)



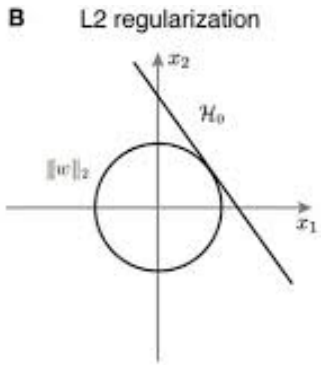
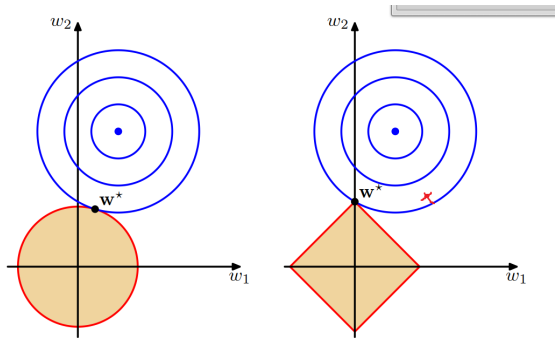
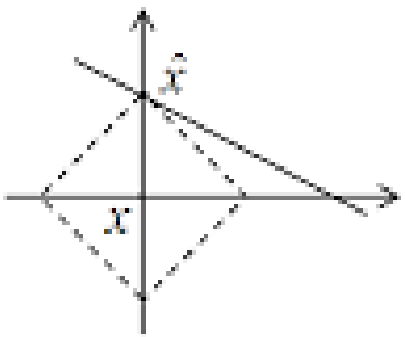
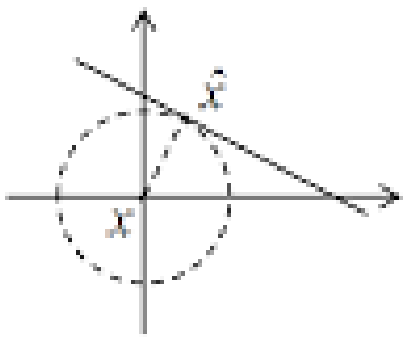
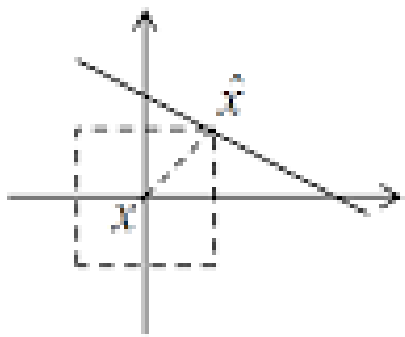


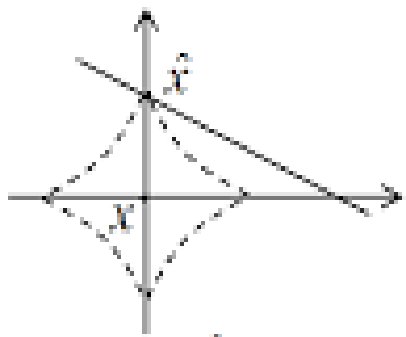
Figure 3.4 Plot of the contours of the unregularized error function (blue) along with the constraint region (3.30) for the quadratic regularizer $q = 2$ on the left and the lasso regularizer $q = 1$ on the right, in which the optimum value for the parameter vector $\mathbf{w}$ is denoted by $\mathbf{w}^\star$.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

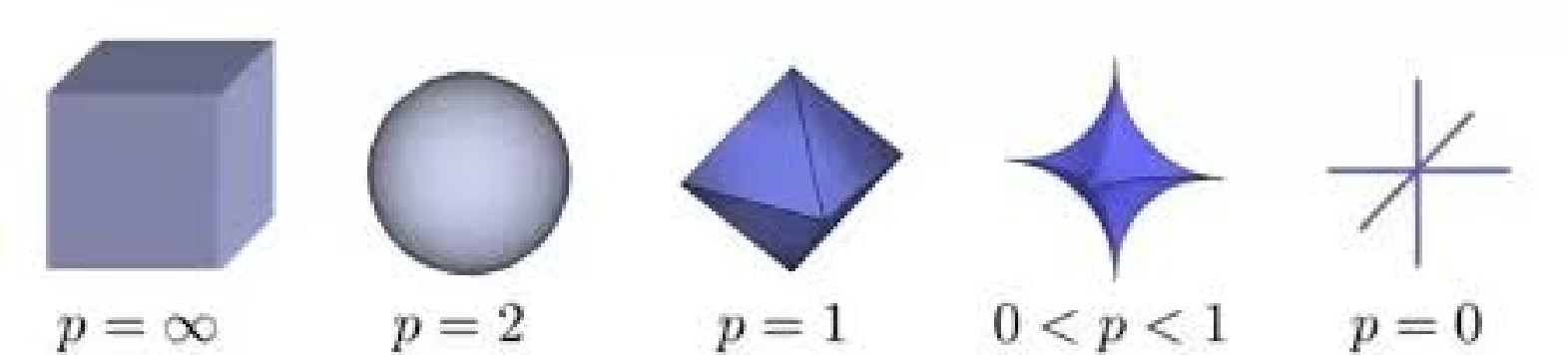












Blank

What Next:?

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