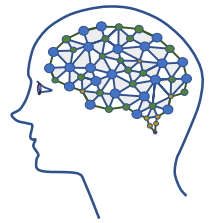
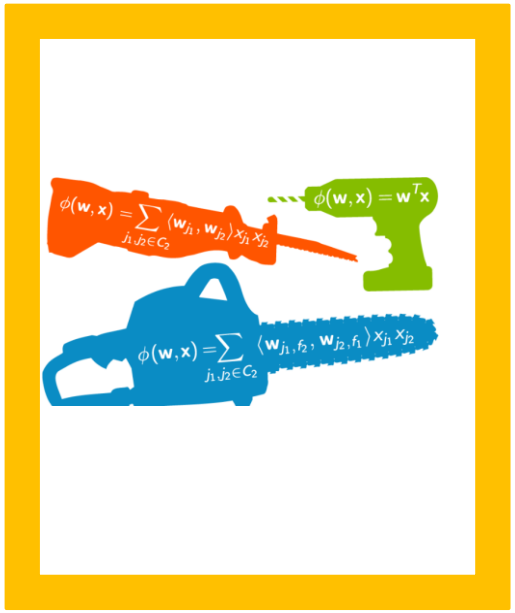
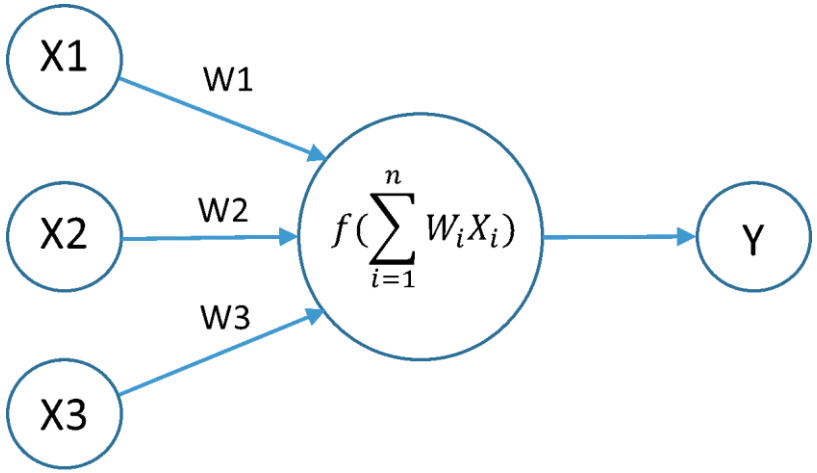


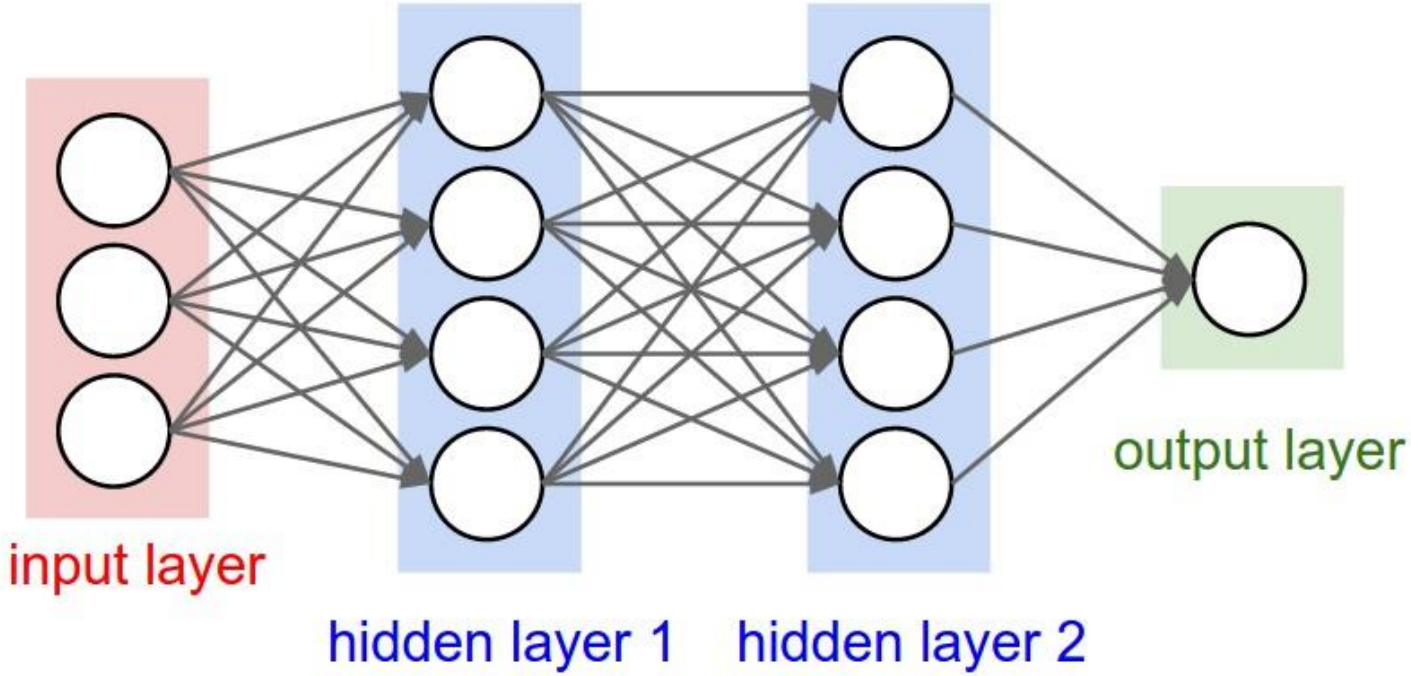


Feature Engineering via Three Examples

Good Features yield Excellent Results.

Review: Neural Networks





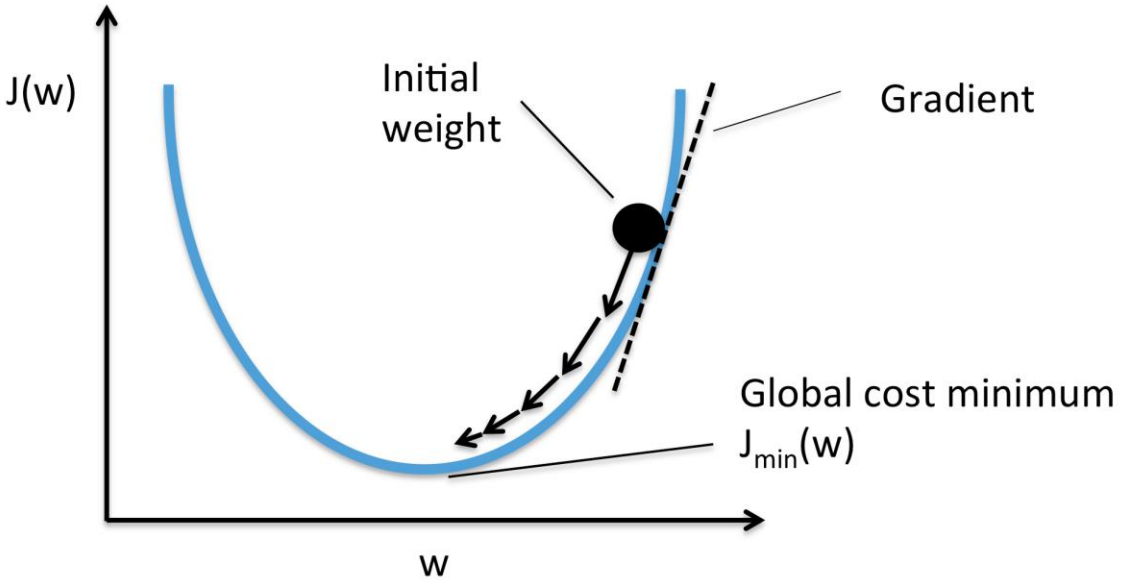
Neuron Model

Y = f($\mathbf{W^T X}$)

f( ) is a nonlinear function

Examples of Deep Neural Networks:
Fully Connected Deep Networks (Fig above)
CNN: Convolutional Neural Networks
RNN: Recurrent Neural Networks
(VGGNet we use today for speech and image is a CNN.)

Review: Gradient Descent



- Popular algorithm for optimization (we will use during training).
- Start with a guess
  - Move on the negative slope
  - Until we reach minima

Review: Pipeline/Concept Map

DATA

Structured (numerical, categorical attributes)
Digital Logs (Tweets, SMS)
RawData/Sensors (Image/Speech)
User behaviors
Etc.

FEATURE

Intuitive User defined
Raw data itself

Statistics (Histograms, PCA)

Signal Process (Fourier Xform)

FEATURE XFORMATIONS

Feature Selection

Feature Extraction

Dimensionality Reduction

Eg. PCA

ML PROBLEMS

1. Classification
   a. Binary
   b. Multiclass
2. Regression
3. Clustering
1. Prediction (time series)

ALGORITHMS

1. KNN
2. Naïve Bayes
3. Perceptron
4. Linear

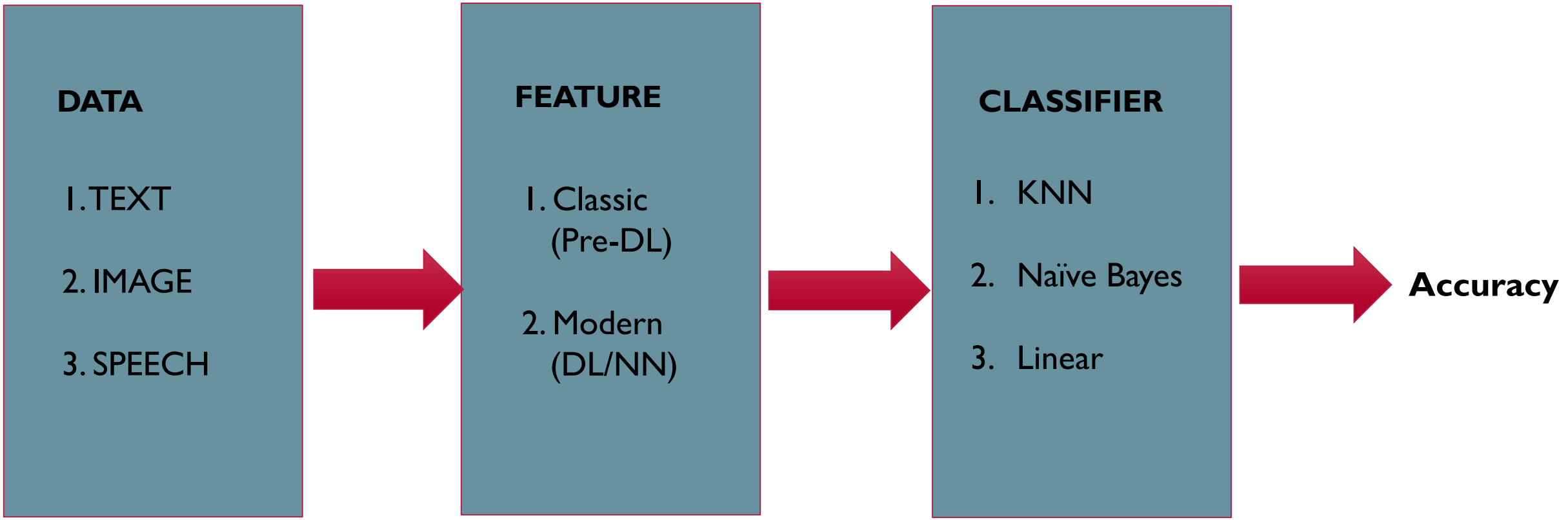
PERFORM. METRICS

Accuracy
Confusion Matrix
Precision
Recall
AP
True Positive
Etc.

Example “Real Problems”

- Spam Filter
- Fraud Identification
- Sentiment Analysis
- Fire Alarm
- Computational Advt.
- Recommender Systems
- Face/Fingerprint Phone Lock
- Stock Price Prediction
- Language Translation
- Speech Recognition
- Product Rating
- Web page Ranking

Plans for Today



Possible to do 18 (=3 X 2 X 3) experiments.

Goal

- Reiterate the Pipeline
- Appreciate that there are many solutions at each stage
  - Best depends on the "exact" problem situation
  - Often best is *too new* for directly understanding; Take steps
- Appreciate the "modern" features/representations
  - **Features that are learned**
  - **Features for complex unstructured data**

Questions?

Agenda

- Representing Text
  - Bag of Words
  - Word2vec
- Comments
- Representing Speech
  - MFCC
  - CNN Features
- Representing Faces
  - Eigen face
  - VGG Face
- Comments:
  - x2vec

Problem Statement 1

- Given a mail-exchange (a long text file), tag the “news group” (such as “hardware”)
- 20 classes (hardware, autos)
- 1000 samples per class
- 950 for training and 50 for testing.

Representations

- A: Bag of Words (Classic)
- B: Word2Vec (Modern)

Bag of Words – Text Domain

- Orderless documentation representation, frequencies of words from a dictionary.

Bag of Words – Text Domain

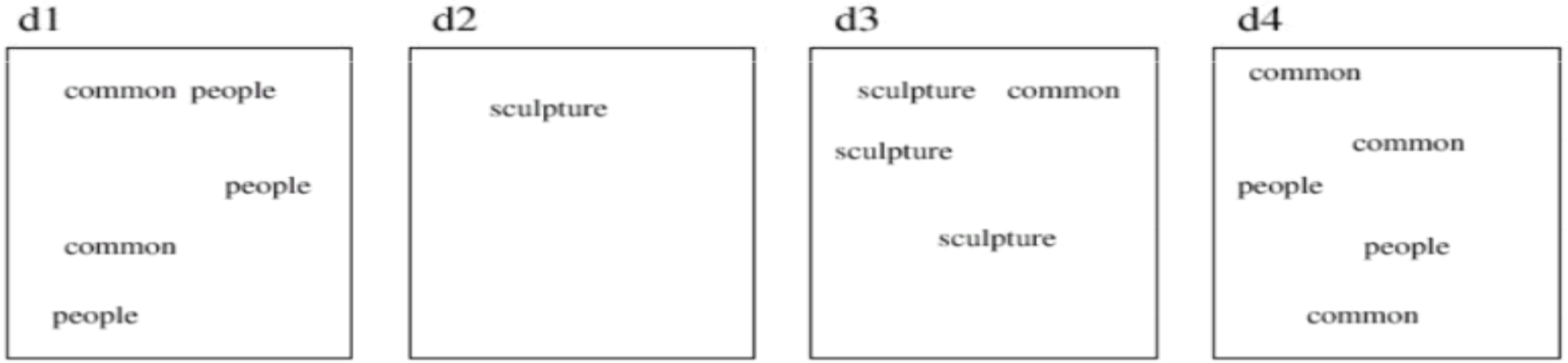
- Orderless documentation representation, frequencies of words from a dictionary.

Bag of Words – Text Domain

- Orderless documentation representation, frequencies of words from a dictionary.

Bag of Words Histogram

- Orderless document representation: frequencies of words from a dictionary.
- Classification to determine document categories.



Bag-of-words

Common	2	0	1	3
People	3	0	0	2
Sculpture	0	1	3	0
...	...	...	...	...

Histogram of Word Occurrences

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
…	…

Comments: Weight Words (eg TF-IDF)

- Not all words are equally useful.
- Stop words: Words that will not add value. (can be removed)
    - Eg. “The” , “of”, “and”
- Frequent words (TF: Term Frequency)
    - Higher the frequency, more useful
- Rare across documents (IDF: Inverse Document Frequency)
- Weight words: Proportional to TF and Inv. Prop to IDF.
- (In the lab, you just remove some frequent and some rare words.)

1-hot representation and Histograms

- Vocabulary sized vector (V)
  - Huge size
- 1 at only one place. Else 0.
- Histogram (BoW) for a documeny
  - Add all such vectors
  - h_i = how many times i^{th} word appear

book [0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0]

library [0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0]

- Disadvantages
  - No semantics
  - Distances have no utility
  - Sparse (lots of zero)
  - High dimensional

Questions?

Word2Vec [Mikolov 2013]

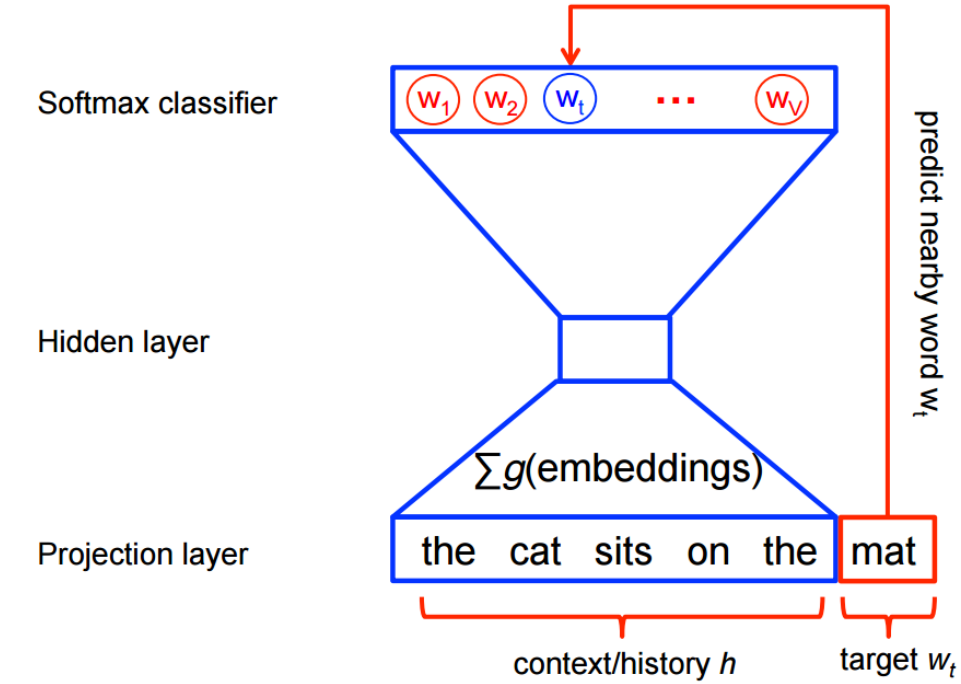
- Predict the missing word from the context (surrounding word). (eg. Fill in the blanks)
- Train a machine learning algorithm

to do this using huge quantity of text

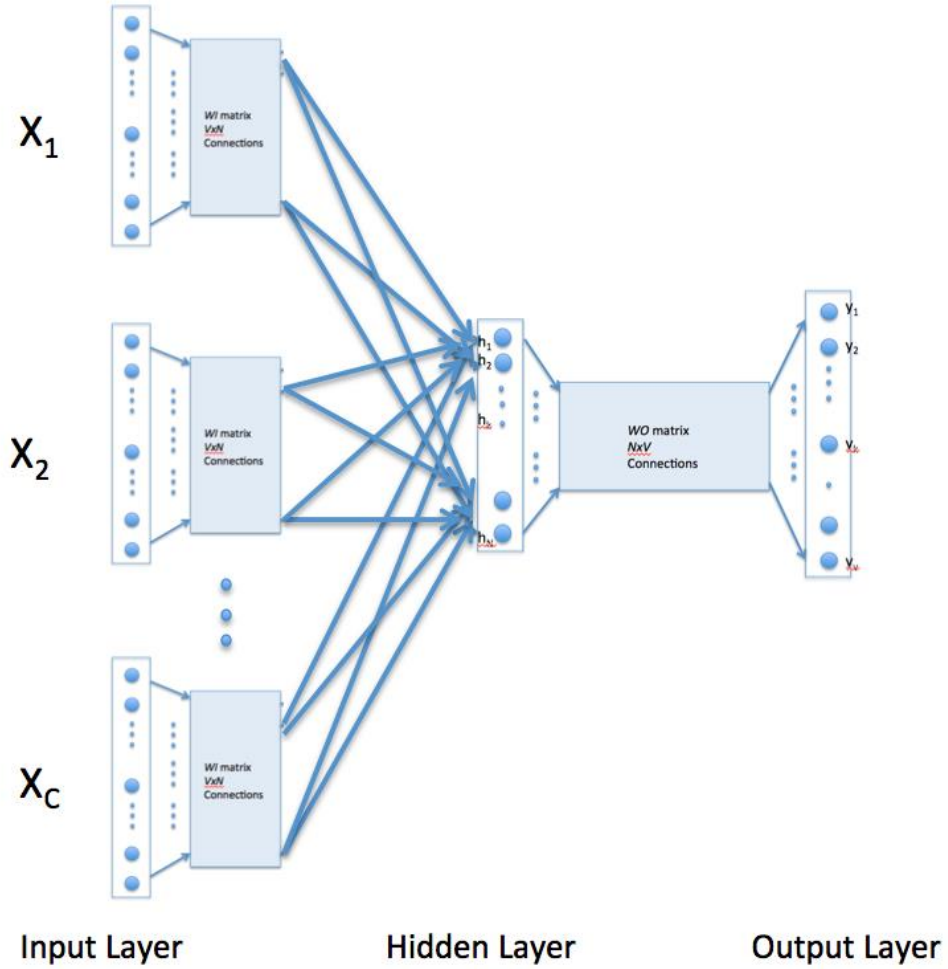
from Internet.

  - Google, Facebook etc.
  - Word2Vec, fastText, Glove

- Available as trained models. (use in the lab)



Word2Vec

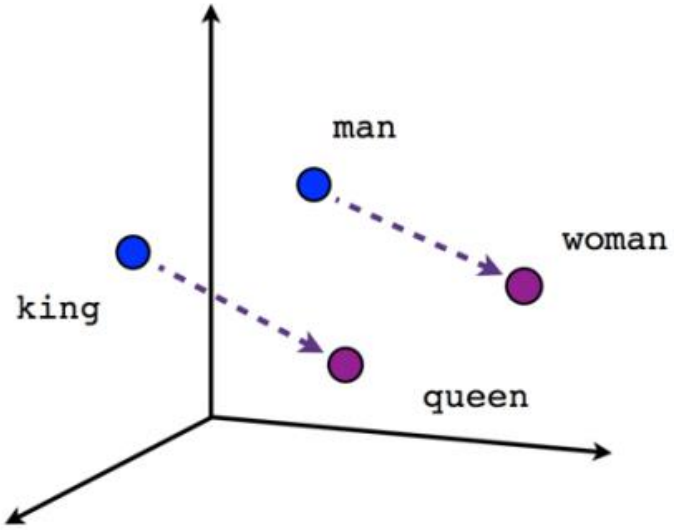


- C-BOW (one of word2vec)
  - **WI**: N X V Matrix
  - **WO**: V X N Matrix
- Input and Output
  - 1-hot representations
- (For lab, N = 300)
- New representation:
  - **WI** $\mathbf{x_i}$

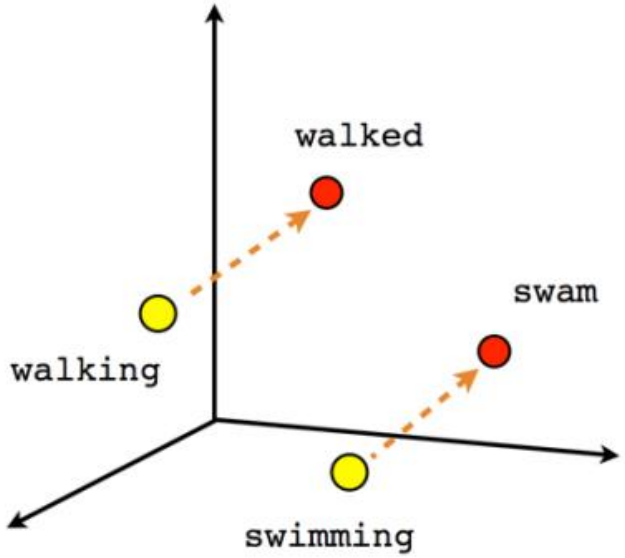
Word2Vec: Comments

- Meaning of a word is defined by the co-occurring words (around).
- Eg.
  - I travelled by Bus. I travelled by Car
- Bus and Car should have similar representations. But not identical. Why?
- Eg.
  - This is our family Car. This is a public Bus.

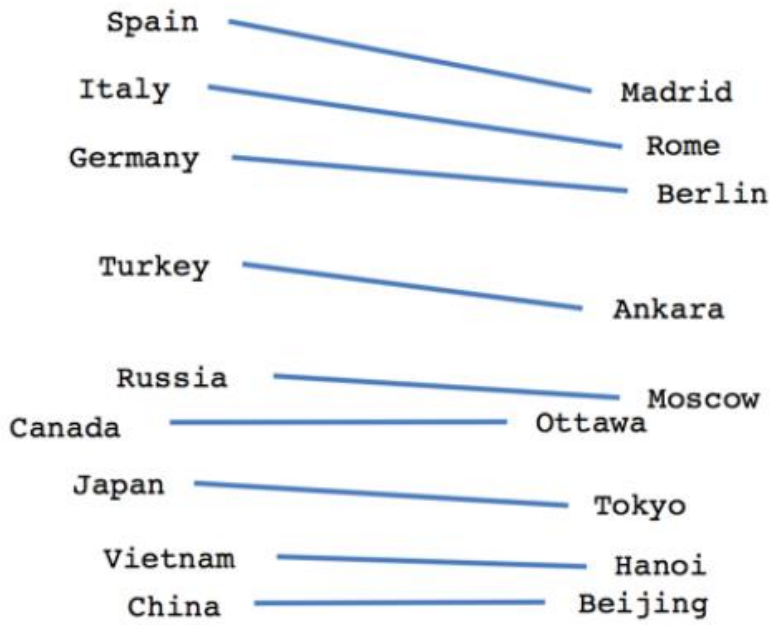
Examples



Male-Female



Verb tense



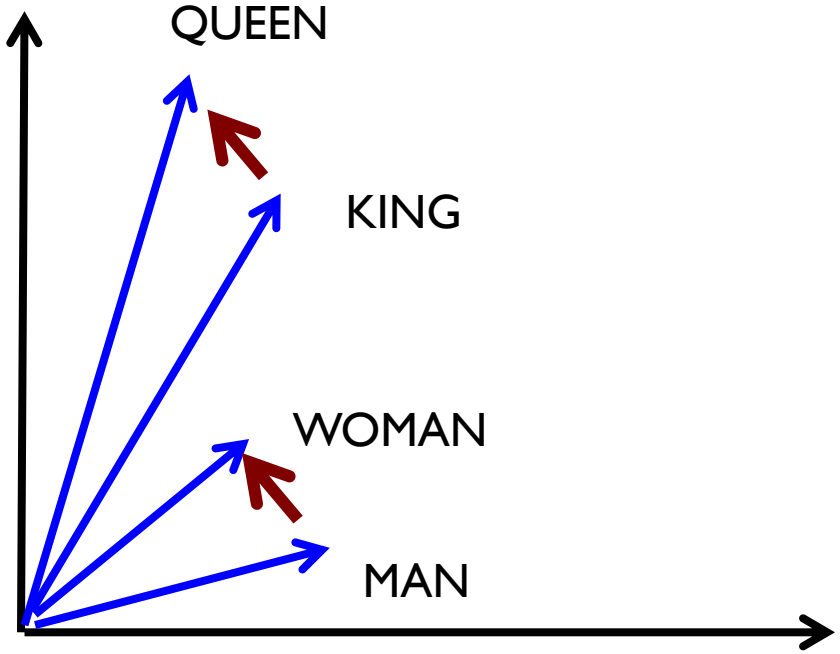
Country-Capital

May read (light reading): https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/

Examples

- I walked 10 Km
- I am walking 10Km
- I swam 10 Km
- I am swimming 10Km

Visualizing “Word Arithmetic”



- Representation for the document can be now the sum of word2vecs.
- (This fancy arithmetic is known earlier also. It became popular with word2vec)

vector[Queen] = vector[King] - vector[Man] + vector[Woman]

Comment: Categorical Variables

- We always worried about "We also have categorical variables".
  - Text is a classical example.
- A 1-hot representation is a possibility
- However, you can have now better dense embedding/representations.
  - Eg. Motivated by word2vec.

Questions?

Problem 2

Step 0: Preprocess to 224 X 224 or 50716 X 1

Recognize Indian Celebrities

10 Classes

20 Example each (15 for training and 5 for testing; No colour.)

Representations

- A Eigen faces (classic)
  - (Same as PCA you already know)
- B: VGG DeepNet Features (modern)
  - (the same you had seen for CFAR)

Eigen Face: Visualization

Mean of All Faces

Top 20 Eigen Vectors (eigen Faces)

Eigen Face Feature

x

$$X = a_1 e_1 + a_2 e_2 + \ldots\ldots + a_{20} e_{20}$$

Any face in the database can be accurately represented as a linear combination of these 20 eigen faces. Representation is the coeff: $a_1, a_2, a_3 \ldots\ldots a_{20}$

Deep VGG Feature

- A deep neural network (CNN) trained only on face for the face recognition task (for 1000(?) classes).
  - 16 Layers deep neural network.
- This is now a generic feature learner for face related task.
  - A generic object recognition network (like from CFAR) could also be used. But with lower performance.
- Feature dimension: 2622

Questions?

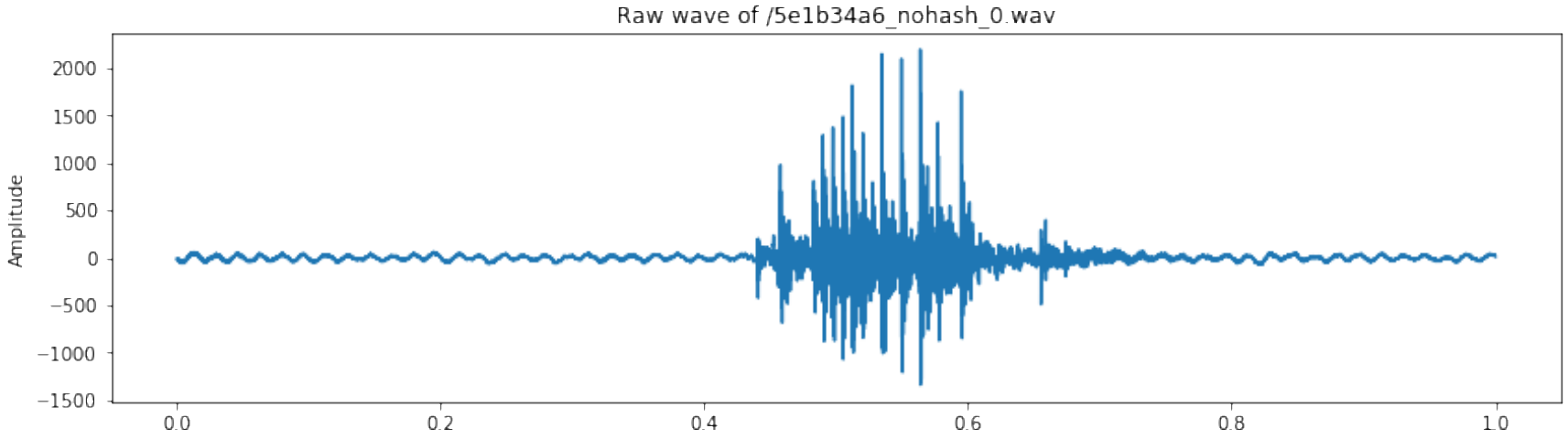
Problem 3

- Sound waves (.wav files)
- 30 short commands (“cat”, “dog”, “go”, “Happy”)
- 1 sec duration
- 65000 samples (many people)

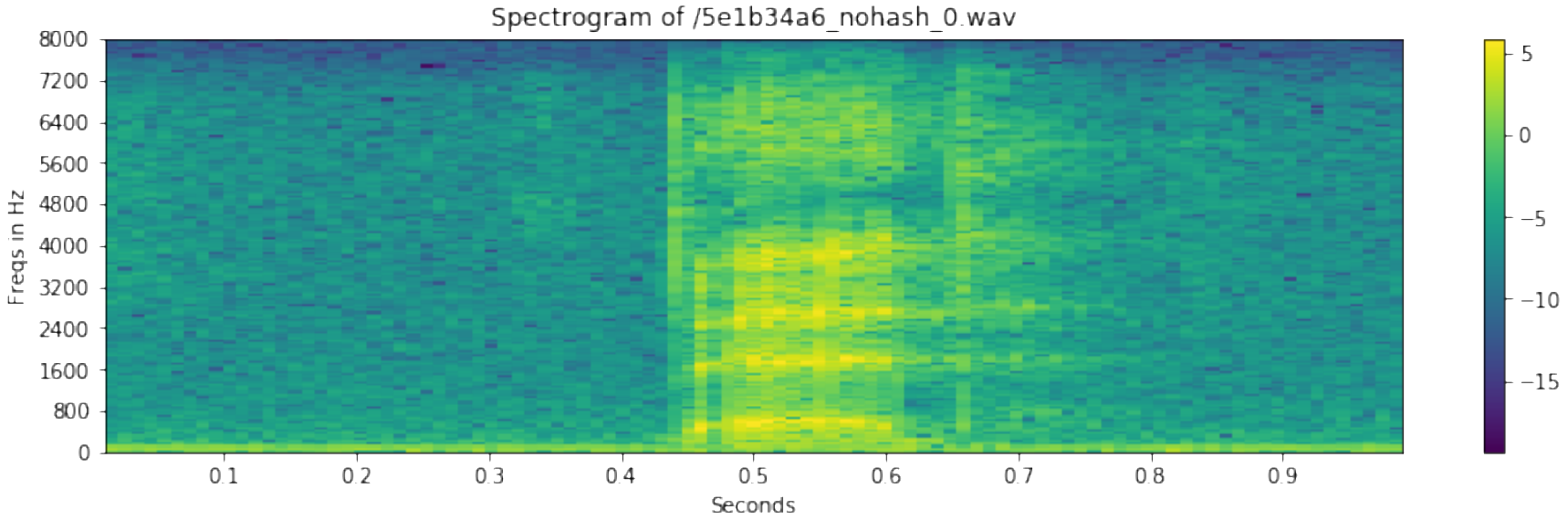
Representations

- A: MFCC (Signal processing based; Classical)
  - Mel Frequency Cepstral Coefficients
- B: CNN based (Modern)
  - VGG Features on the Mel Spectrogram

Classical Feature (MFCC)

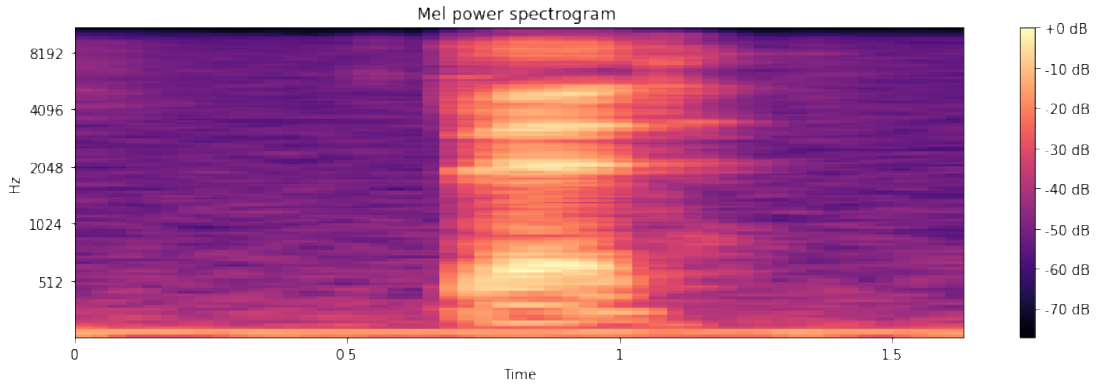


Amplitude Vs Time



Frequency Vs Time

Features from Mel Spectrogram



MFCC
(Hand coded Classic Features)

VGG19-Features
(Trained on Mel spectrograms)

Some questions

- How do I use/extend in my domain?
  - Every problem has its own characteristics. Let us not assume that we can copy solutions.
  - There are numerous examples of generic notions getting employed or extended to your notion. (please see next slide)

Representing Complex Entities

- Products:
  - Product has attributes (price, manufacture), Opinions/Review, Product Images, Product description
- **Represent items (and sometimes users) as vectors in the same space and use their distances to compute recommendations.**
- Many recent works:
  - Prod2Vec: Consider in a sequence the products that a customer buy.
  - MetaProd2Vec: Also bring in meta data (categories, brands, tags etc.)

Representing Complex Concepts

- Medical Records
  - Visits, Numerical attributes, Personal attributes, Images, Reports
- Data:
  - Visits: A sequence of visits of patients
  - Medical terms (diagnosis, disease, medication, procedure)
- See Med2Vec [2016]

“X”2Vec

- Motivated by word2vec. More examples.
  - Node2vec (network nodes, social networks)
  - Doc2vec (eg. Paragraphs, larger documents)
- Many More:
  - See: https://github.com/MaxwellRebo/awesome-2vec
  - Tweet2vec, wiki2vec, item2vec, author2vec, playlist2vec, game2vec,
- Learn from domain examples.
- An embedding that can be manipulated.

Summary

- Features are integral part of the ML solution
  - Good features take us a long distance.
  - Intuitions are important. Data is more important.
  - If you can learn, that is still better.
- Arrival of the new category of features due to DL
  - Huge performance improvements in many areas.
- Domains may have novel ideas on defining, learning and using representations.

Thanks!!. Questions?

Train, Val and Test

- Ideally Data needs to be spit into three (not two).
  - Training (Train), Validation (Val) and Testing (Test).
- You do not see test data. Use only once at the end.
  - This is like your examination data.
- Training is split into two parts: Train and Val.
  - Val is used to find the best "hyper parameters" (like K in K Means). Val is used indirectly in the design. Therefore, it is not really the the test.
- In labs, we often use train and test (here, val and test are same). You may be even tuning the parameters for best results on val/test here.