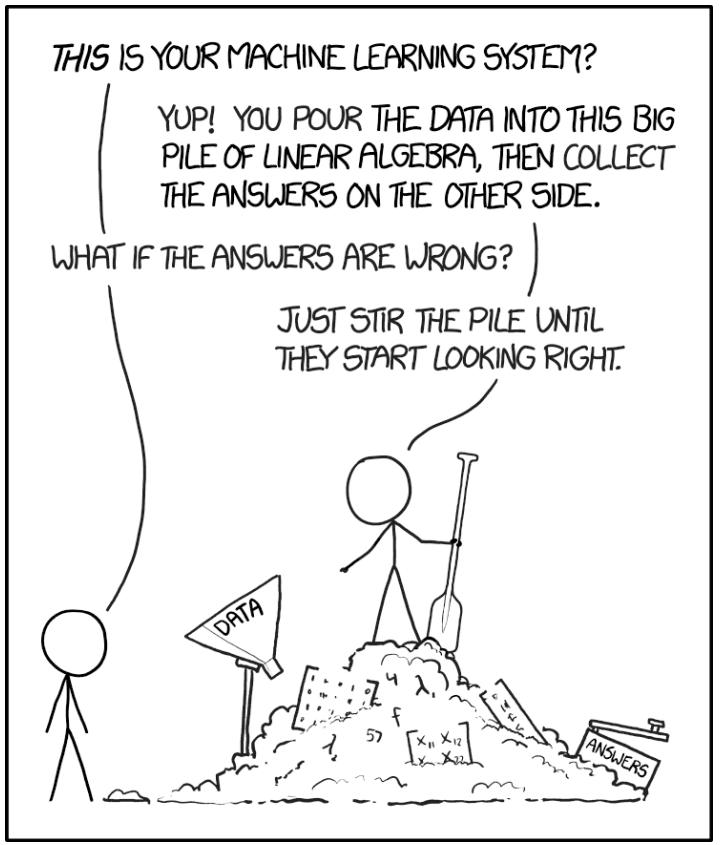
COMP 4446 / 5046
Lecture 2: Foundation of NLP Systems

Jonathan K. Kummerfeld

Semester 1, 2025

THE UNIVERSITY OF SYDNEY

Machine Learning




```
[The pile gets soaked with data and starts to get
mushy over time, so it's technically recurrent.]
```

Source: https://xkcd.com/1838/

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Assignment 1

Due: Tuesday of week 3 (11:59pm)

Collaboration rules:

- Individual
- You may discuss the questions with others in the class, but you should write the code yourself
- Do not share the assignment questions online or with people outside the course
- Automatic plagiarism detection will be run on all submissions
- If you use AI assistance, document what you ask and get back and include it in a text file that you submit with your answers.

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Marks

Now visible on Canvas on the “Marks Portal” page.

If something seems incorrect, submit this form (link also on the Assessments information page in Canvas):

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Core Components of NLP Systems

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

NLP systems have five key components

Data	Examples of the language phenomena we want our system to handle
Model	A function that maps (input, output) pairs to scores
Inference Method	A way to make a prediction for an example given a Model
Metric	A function that gives a score to the output produced by a Model given some Data
Learning Method	A way to update a Model given Data a Metric and an Inference Method

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We'll use sentiment analysis as a running example

This movie was great! Would watch again	+
The movie was gross and overwrought, but I liked it	+
This movie was not really very enjoyable	-

Example from Bo Pang, Lillian Lee, Shivakumar Vaithyanathan (2002),
via Greg Durrett's Lecture Slides for CS378 at UT Austin

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Here are these core components for **rule based** methods

Data	Not strictly necessary, but rules are hard to write without examples
Model	A set of rules
Inference Method	Code to run the rules
Metric	0 if the prediction matches the true answer, 1 otherwise
Learning Method	People edit the rules, or use a rule finding algorithm

If the review contains "good" return positive

```
for rule in rules:
  if rule(input):
    return true
```

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

2

Here are these core components for **linear** models

Data	A set of examples with their true labels	*This movie was great! Would watch again* + *The movie was gross and overwrought, but I liked it* + *This movie was not really very enjoyable* -
Model	A set of features and weights	*"liked": +1.2,* *"gross": -1.4,* *"gross AND liked": +0.3, …*
Inference Method	A range of methods	`prediction = sign[sum of features in the review]`
Metric	A range of methods	`1 if correct else 0`

Learning Method	A range of methods	`if prediction is wrong:`
		`  for each feature used:`
		`  feature += sign of answer`

COMP 4446 / 5046
Lecture 2, 2025

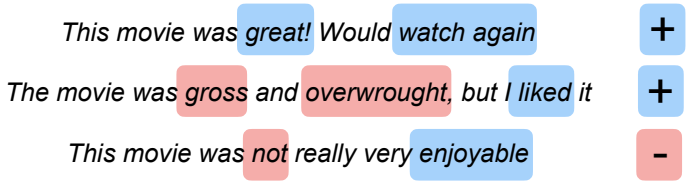
Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

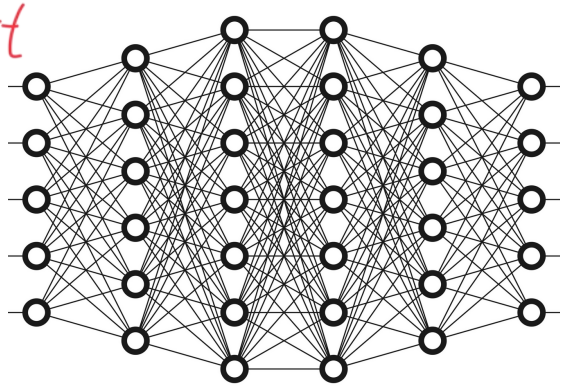
3

Here are these core components for **neural networks**

Data	A set of examples with their true labels
Model	A set of weights and structure of the neural network
Inference Method	
Metric	A range of methods
Learning Method	A range of methods



more next week



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What can data look like?

Data

Model

Inference Method

Metric

Learning Method

(input, output)

Input - Various sized pieces of text

Output	**Example**
A set of labels	(sentence, sentiment)
Structured	(transcript, reply-to links)
Free text	(document, summary)

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

How much data do we need?

Data

Model

Inference Method

Metric

Learning Method

Sentiment:
12,000 sentences

Sentence Structure:
40,000 sentences

Translation:
60 million words from each of 21 languages

Language modelling:
1 million - 30 trillion+ words

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What does a model look like in code?

Data

Model

Inference Method

Metric

Learning Method

receive input

```
def model(text, label):
    . . .
    do stuff
    . . .
    . . .
    return score
```

return output

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What does a model look like in maths?

Data

Model

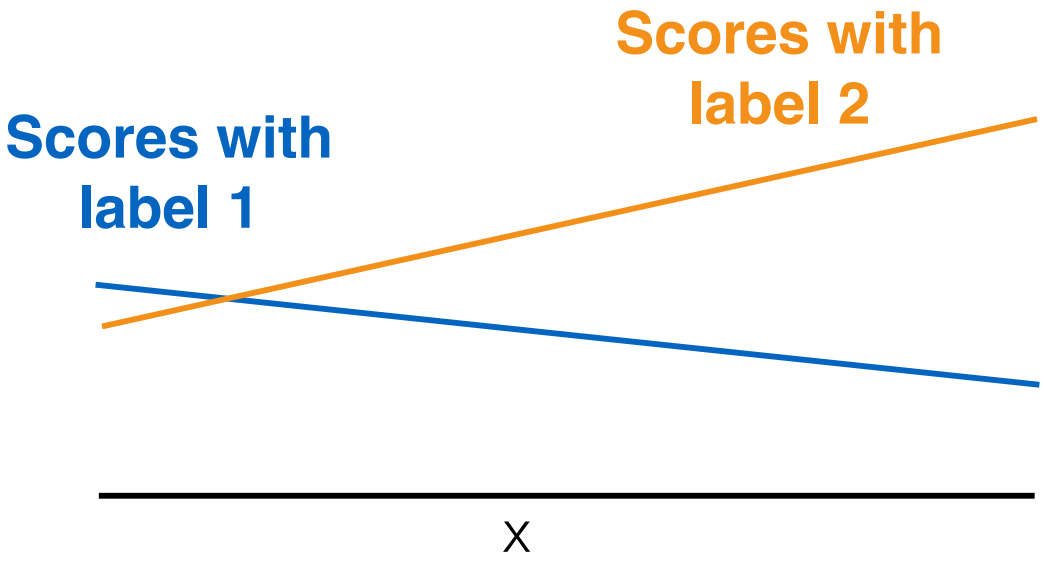
Inference Method

Metric

Learning Method

$$\text{score}_1 = a_1 x + k_1$$

$$\text{score}_2 = a_2 x + k_2$$



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What does a model look like in maths?

Data

Model

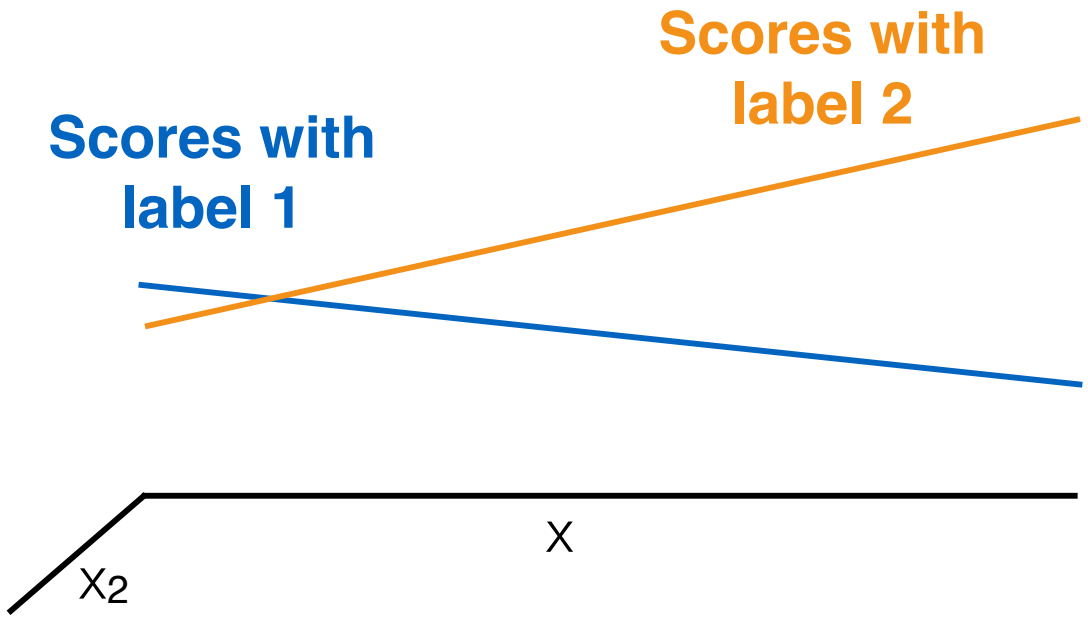
Inference Method

Metric

Learning Method

$$\text{score}_1 = a_1 x_1 + b_1 x_2 + k_1$$

$$\text{score}_2 = a_2 x_1 + b_2 x_2 + k_2$$



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What does a model look like in maths?

Data

Model

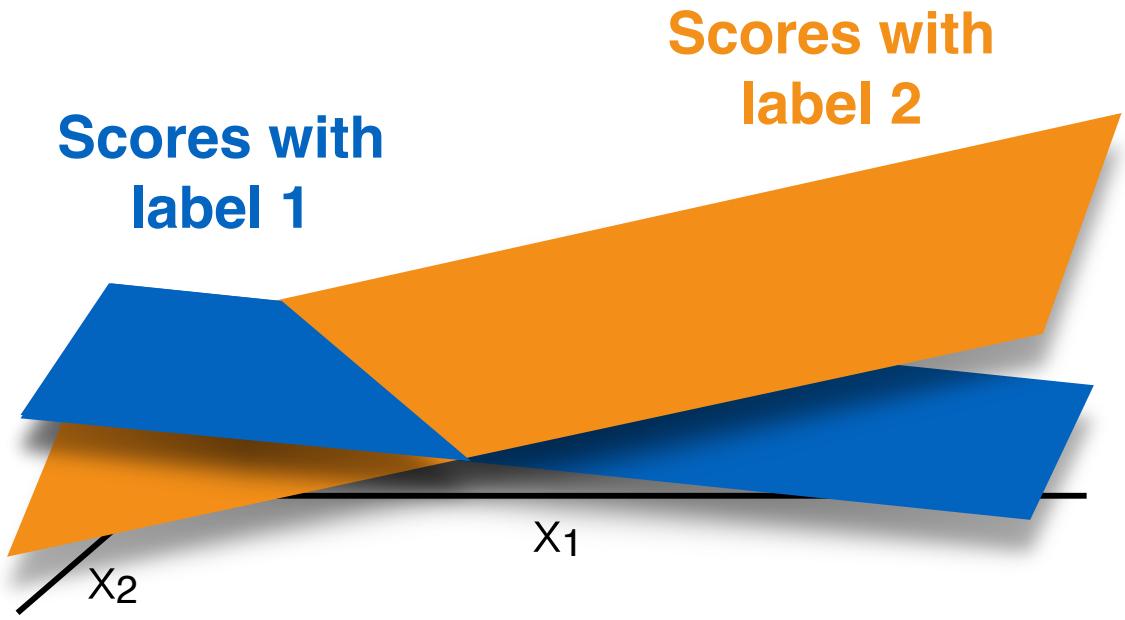
Inference Method

Metric

Learning Method

$$\textbf{score}_1 = \mathbf{a_1}x_1 + \mathbf{b_1}x_2 + \mathbf{k_1}$$

$$\textbf{score}_2 = \mathbf{a_2}x_1 + \mathbf{b_2}x_2 + \mathbf{k_2}$$



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What does inference look like?

Data

Model

Inference Method

Metric

Learning Method

Simple approach:

```
for each possible output:
  calculate score
  If score is better than best so far
    update best
```

What if there are lots of possible outputs?

Translation:
all possible sentences

Answer span retrieval:
n^2

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What are some algorithms for inference?

Data

Model

Inference Method

Metric

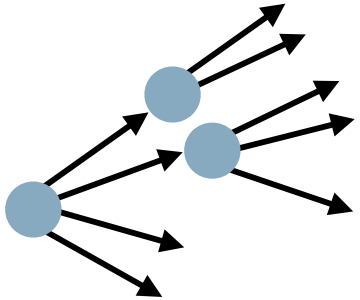
Learning Method

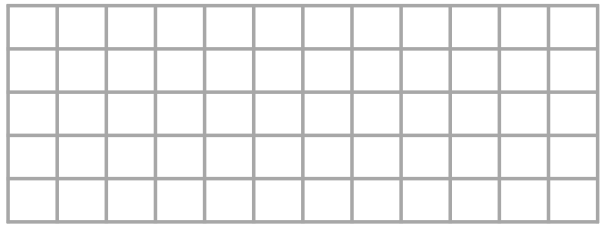
Greedy Search (1)

A* Search (2)

Dynamic Programming (3)

Best($word_1$)
Best($word_2$ given $word_1$)
…





COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What are some algorithms for inference?

Data

Model

Inference Method

Metric

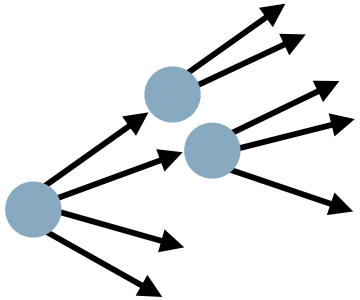
Learning Method

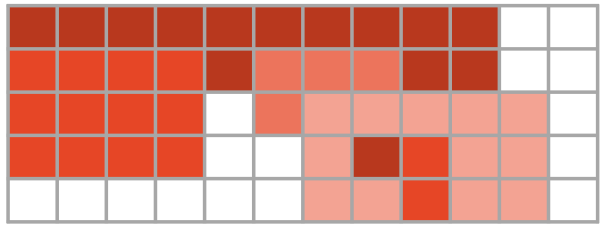
Greedy Search

A* Search

Dynamic Programming

Best($word_1$)
Best($word_2$ given $word_1$)
…





COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What are some metrics?

Data

Model

Inference Method

Metric

Learning Method

y	True label
x	Input
$f(x)$	Function output
$p(y_i)$	Probability of specific output

0-1 Loss

$$\begin{cases} 0 & \text{if correct} \\ 1 & \text{otherwise} \end{cases}$$

Hinge Loss penalize wrong predict

$$\max(0, 1 - y * f(x))$$

Cross-Entropy Loss

$$-\log(p(y))$$

$$= -\sum_{i \in \text{classes}} y_i \log(p(y_i))$$

Squared Error

$$(y - f(x))^2$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What is the intuition for learning?

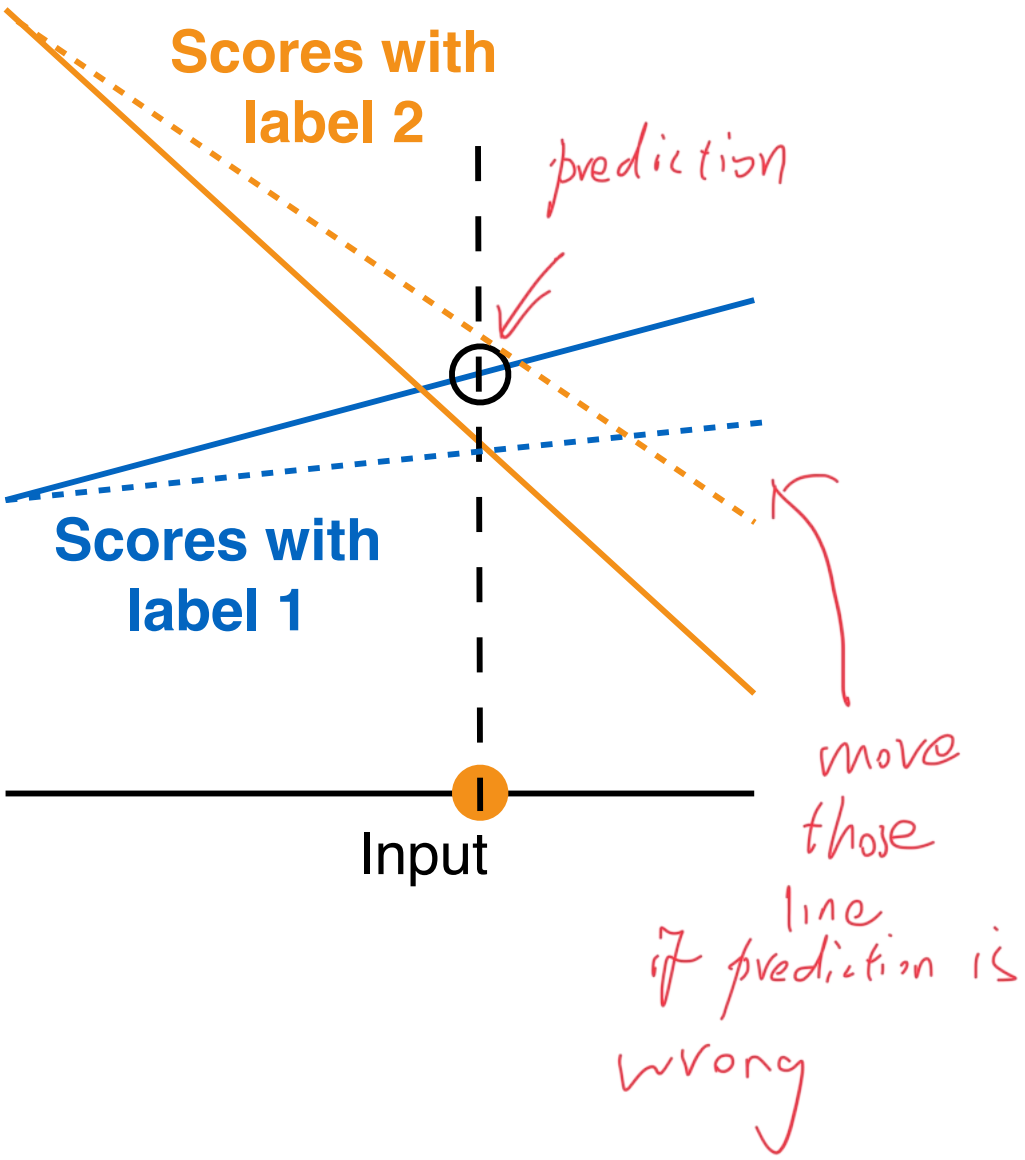
Data

Model

Inference Method

Metric

Learning Method



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What is the intuition for learning?

Data

Model

Inference Method

Metric

Learning Method

$$score_1 = \mathbf{a_1}\, x_1 + \mathbf{b_1}\, x_2 + \mathbf{...}$$

How many examples at once?
1, a few, all

How much should **a, b**, etc change?
+/-1, vary based on error, vary based on previous updates

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

The Baseline: Supervised, Greedy, and Linear

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

NLP systems have five key components

Data	Examples of the language phenomena we want our system to handle
Model	A function that maps (input, output) pairs to scores
Inference Method	A way to make a prediction for an example given a Model
Metric	A function that gives a score to the output produced by a Model given some Data
Learning Method	A way to update a Model given Data a Metric and an Inference Method

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Examples of classification problems

Data

Spam email detection	Spam / Normal
Authorship attribution	Hamilton / Madison / Jay / Other
Legal responsiveness	Responsive / Not
Sentiment	Positive / Negative
Demographic prediction e.g., Gender, Age, Location, …	

Ethical concerns?

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Input is a document and output is a label

Data

Input

Output

- ☐ Option 1
- ☐ Option 2
- ☑ Option 3
- ☐ Option 4

Various terms: option, label, class, category, …

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Data

Tokenisation

Simplest - split on whitespace

That's pretty good in English!

However…

並非所有語言都有空格

(Not all languages have spaces)

Der Abfallwirtschaftsinspektor der Universität

(The university waste management inspector)

More on this later in the course

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Supervised learning means we use labeled data

Data

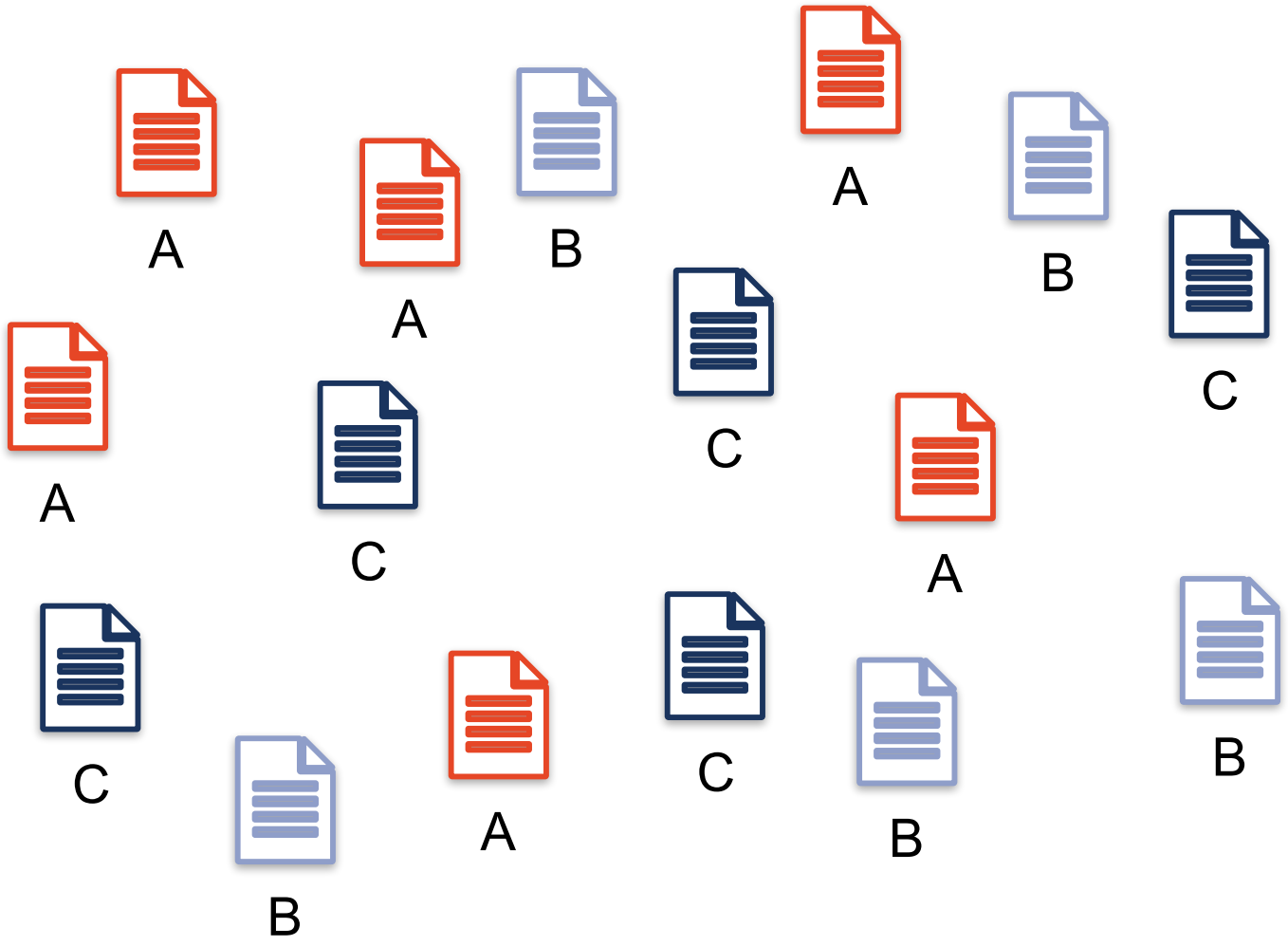
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Supervised learning means we use labeled data

Data



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

NLP systems have five key components

Data	Examples of the language phenomena we want our system to handle
Model	A function that maps (input, output) pairs to scores
Inference Method	A way to make a prediction for an example given a Model
Metric	A function that gives a score to the output produced by a Model given some Data
Learning Method	A way to update a Model given Data a Metric and an Inference Method

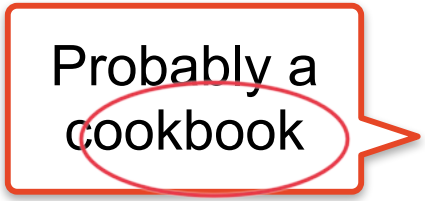
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Bag of Words (BoW) representation

Model



biscuit	10	0
brownie	5	0
cake	50	0
chocolate	143	1
flan	2	0
mousse	3	0
muffin	14	0
tart	27	0
the	1,152	2,483
…	…	…

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's estimate storage costs for different methods

Model

1-Hot Vector

100,000 words in vocabulary
64 bit integer counts
1,000 documents

763 mb!

Assume each document is one sentence, ~25 words.
At most 25 counts are non-zero!

[0, 0, 0, 1, 0, 0, 0 … 1 … 2 …]

Dictionary

Python: ~2 mb

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

BoW does not account for word senses

Model

I have money at the **bank**. n.

I climbed out of the river on to the **bank**.

I will **bank** that check later today. V.

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Unknown words are ignored in the bag

Model

Test data includes:

That will be $1989.22 please

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Unknown words are ignored in the bag

Model

Test data includes:

That will be $1989.22 please

Unknown word / token isn't counted

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

What do we do with common words like ‘the’?

Model

“Stopwords"

Information Retrieval
- remove/ignore them (in early work)

NLP
- makes very little difference generally
- very important sometimes (e.g. authorship attribution)

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Creative ‘words’ - features

Model

We can put anything in the bag!

Bigrams:

“This chocolate is absolutely delicious”

This_chocolate
chocolate_is
is_absolutely
absolutely_delicious

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Creative ‘words’ - features

Model

We can put anything in the bag!

What to do about negation?

didn’t like this movie , but I

Modify words (here, adding “NOT_” up until the next punctuation symbol):

didn’t NOT_like NOT_this NOT_movie , but I

I didn’t not like this movie

还可以通过加上NOT来反之.

Example from Bo Pang, Lillian Lee, Shivakumar Vaithyanathan (2002), via Dan Jurafsky’s Lecture Slides at Stanford

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Creative ‘words’ - features

Model

We can put anything in the bag!

Lexicons
e.g., list of city names, list of employee names

```
if word is in city_name_list:
  put “city_name_present” in the bag
```

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We combine our counts with weights to get a score

Model

Counts: [2, 0, 0, …]

Weights: [0.2, 1.833, 0.005, …]

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We often include a bias weight, to favour one label

Model

Counts: [2, 0, 0, …]

Weights: [0.2, 1.833, 0.005, …]

$$\text{score} = \sum_{f \in \text{features}} \text{count}_f * \text{weight}_f$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We often include a bias weight, to favour one label

Model

Counts: [2, 0, 0, … , 1]

Weights: [0.2, 1.833, 0.005, …]

$$\text{score} = \sum_{f \in \text{features}} \text{count}_f * \text{weight}_f$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We often include a bias weight, to favour one label

Model

Counts: [2, 0, 0, … , 1]

Weights: [0.2, 1.833, 0.005, … , 0.03]

$$\text{score} = \sum_{f \in \text{features}} \text{count}_f * \ldots$$

We'll discuss where these come from when we get to the Learning Method

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

NLP systems have five key components

Data	Examples of the language phenomena we want our system to handle ✓
Model	A function that maps (input, output) pairs to scores ✓
Inference Method	A way to make a prediction for an example given a Model
Metric	A function that gives a score to the output produced by a Model given some Data
Learning Method	A way to update a Model given Data a Metric and an Inference Method

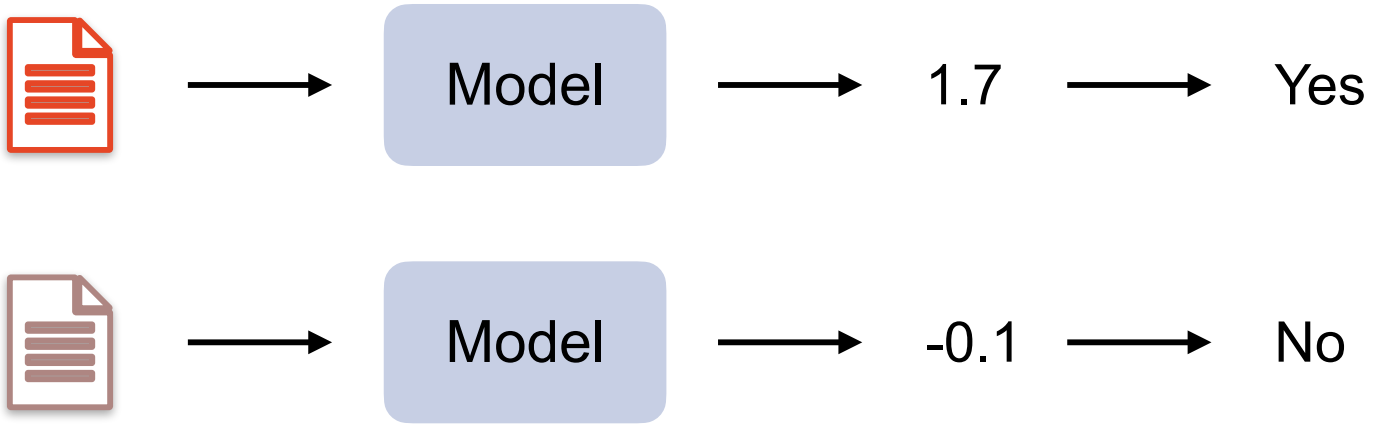
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

For binary classification, we can check the sign

Inference Method



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

For more classes, use several sets of weights

Model

Counts: [2, 0, …]

Weights: [0.2, 1.833, …] score = $\sum_{f \in \text{features}} \text{count}_f * \text{weight}_f$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

For more classes, use several sets of weights

Model

Counts: [2, 0, …]

model just give score

Weights$_A$: [0.2, 1.833, …] $\text{score}_A = \sum_{f \in \text{features}} \text{count}_f * \text{weight}_{fA}$

Weights$_B$: [1.2, -0.4, …] $\text{score}_B = \sum_{f \in \text{features}} \text{count}_f * \text{weight}_{fB}$

Weights$_C$: [0.01, 3.7, …] $\text{score}_C = \sum_{f \in \text{features}} \text{count}_f * \text{weight}_{fC}$

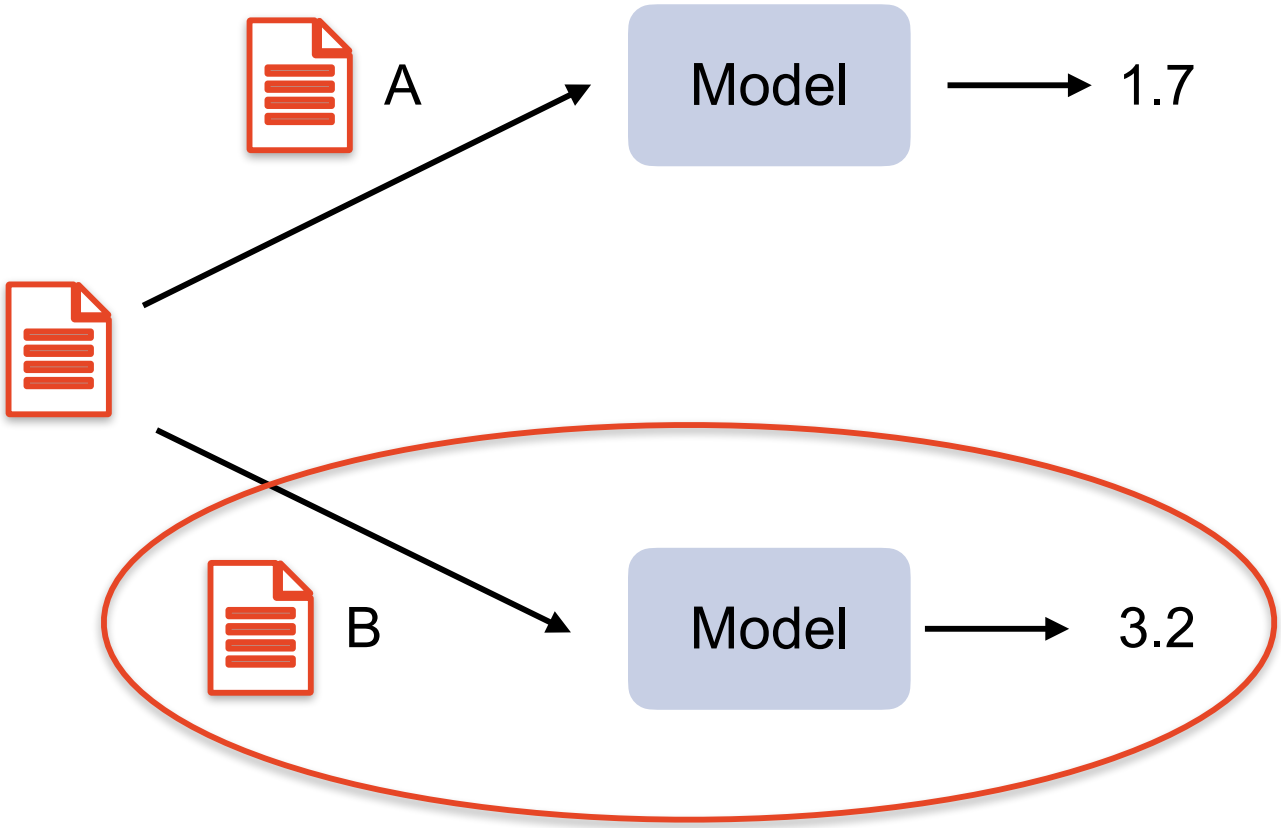
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

For multiple classes, we can compare them all

Inference Method



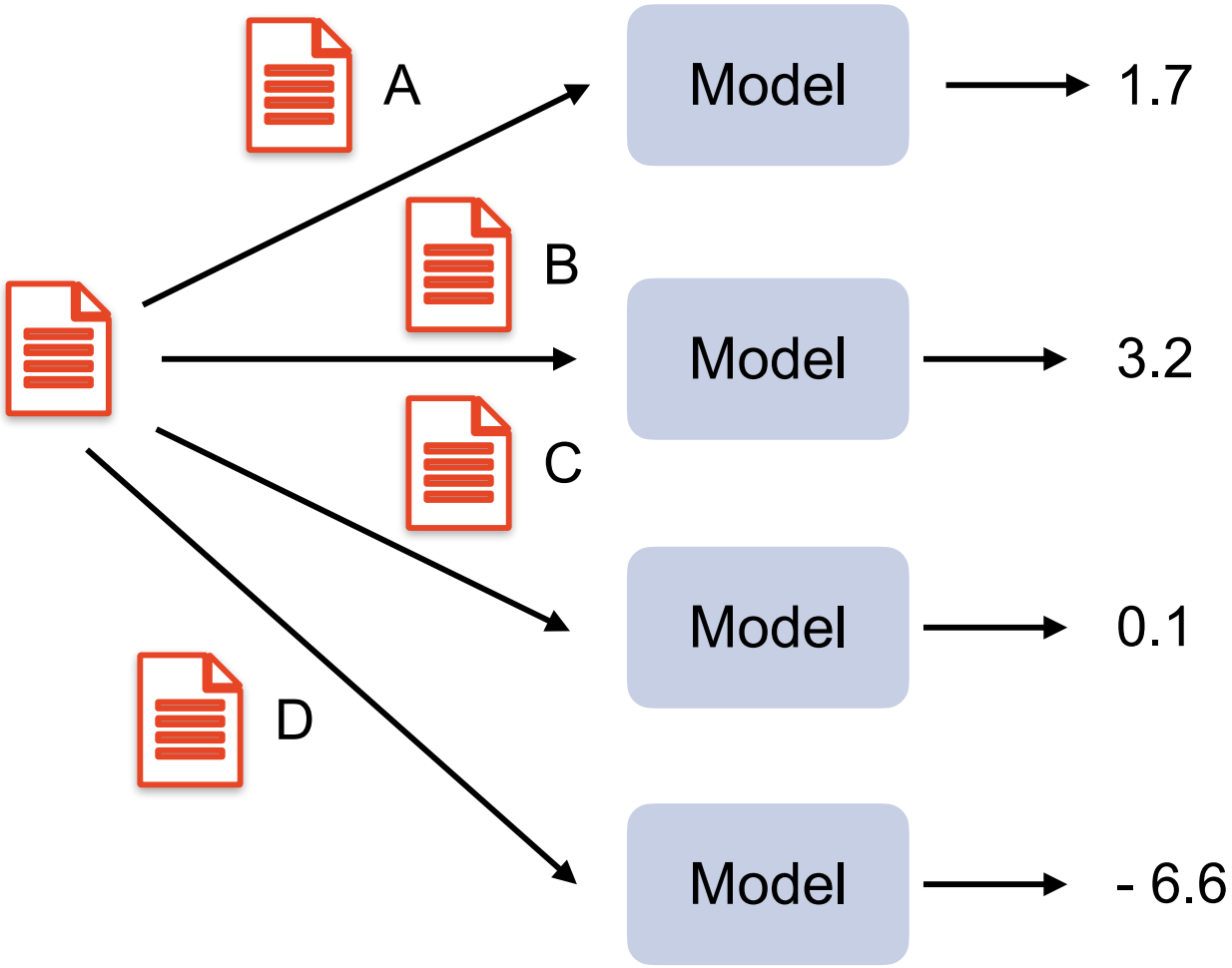
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

For a small number of classes, we can test them all

Inference Method



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

For a small number of classes, we can test them all

Inference Method

```
best = (-infinity, None)
for label in labels:
    score = model(data, label)
    If score > best[0]:
        best = (score, label)

return best[1]
```

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We can express these in vector form

Inference Method

Model

$$\text{label}(\text{doc}) = \underset{l \in \text{labels}}{\text{argmax}}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We can express these in vector form

Inference Method

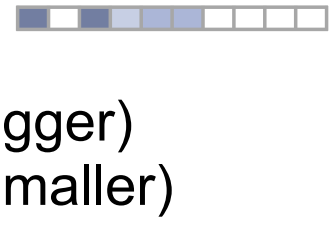
Model

$$\text{label(doc)} = \underset{l \in \text{labels}}{\text{argmax}} \ \text{features(doc)} \cdot \text{weights}_l$$

model

$$\text{label(doc)} = \text{max-index}(\text{features(doc)} \cdot \text{weights})$$

$$= \text{max-index}\left(\quad \cdot \quad \right)$$



Blue = Positive (darker is bigger)
Red = Negative (darker is smaller)

$$= \text{max-index}\left(\quad \right)$$

$$= 3$$

choose the biggest score one inference

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We can express these in vector form

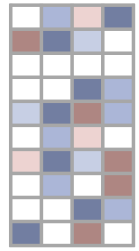
Inference Method

Model

$$\text{label}(\text{doc}) = \underset{l \in \text{labels}}{\text{argmax}} \ \text{features}(\text{doc}) \cdot \text{weights}_l$$

$$\text{label}(\text{doc}) = \text{max-index}\left(\text{features}(\text{doc}) \cdot \text{weights}\right)$$

$$= \text{max-index}\left(\quad \cdot \quad \right)$$



Use a sparse vector!

$$= \text{max-index}\left(\quad \right)$$

$$= 3$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

NLP systems have five key components

Data	Examples of the language phenomena we want our system to handle ✓
Model	A function that maps (input, output) pairs to scores ✓
Inference Method	A way to make a prediction for an example given a Model ✓
Metric	A function that gives a score to the output produced by a Model given some Data
Learning Method	A way to update a Model given Data a Metric and an Inference Method

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

The metric and learning method are closely related

Metric

0-1 Loss

$$\begin{cases} 0 & \text{if correct} \\ 1 & \text{otherwise} \end{cases}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

NLP systems have five key components

Data	Examples of the language phenomena we want our system to handle ✓
Model	A function that maps (input, output) pairs to scores ✓
Inference Method	A way to make a prediction for an example given a Model ✓
Metric	A function that gives a score to the output produced by a Model given some Data ✓
Learning Method	A way to update a Model given Data a Metric and an Inference Method

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

How do we choose these weights?

Learning Method

Two general approaches:

(1) Carefully define a probabilistic model and estimate it, e.g., for Multinomial Naive Bayes:

$$\text{weight}_{f,c} = \frac{\text{count}(f \text{ in docs with class } c)}{\text{count}(\text{all features in docs with class } c)}$$

(2) Start with random values, iteratively improve, e.g., Perceptron

COMP 4446 / 5046
Lecture 2, 2025

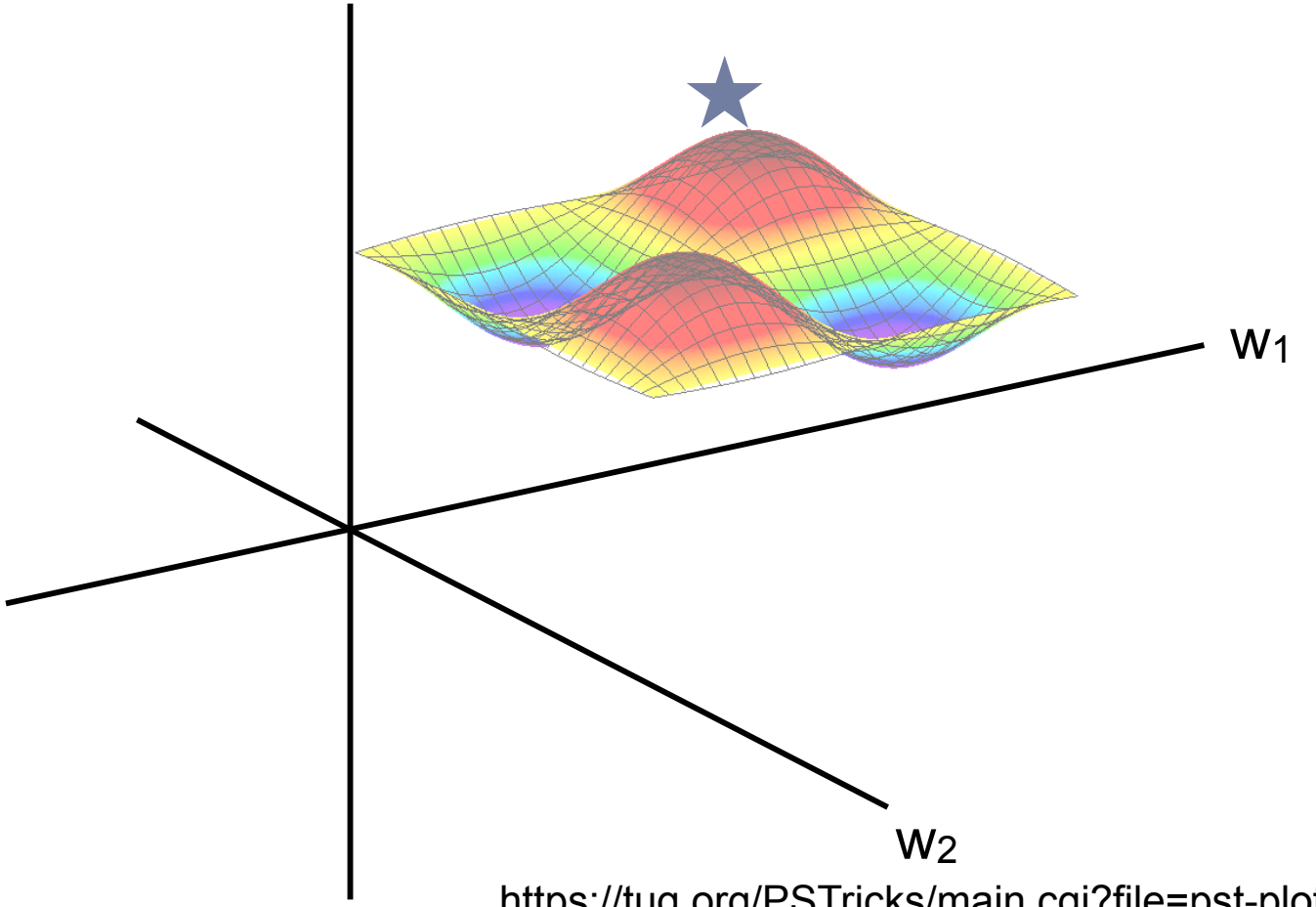
Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

How do we choose these weights?

Learning Method

y = Loss, how bad our errors are



https://tug.org/PSTricks/main.cgi?file=pst-plot/3D/contour

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

One simple approach is the Perceptron

Learning Method

Get the features using our **Data** representation

Get the best label according to our **Model**

If it's right, do nothing
If it's wrong, update our **Model** by:

Adding 1 to the weights for the true label
Subtracting 1 from the weights for the predicted label

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

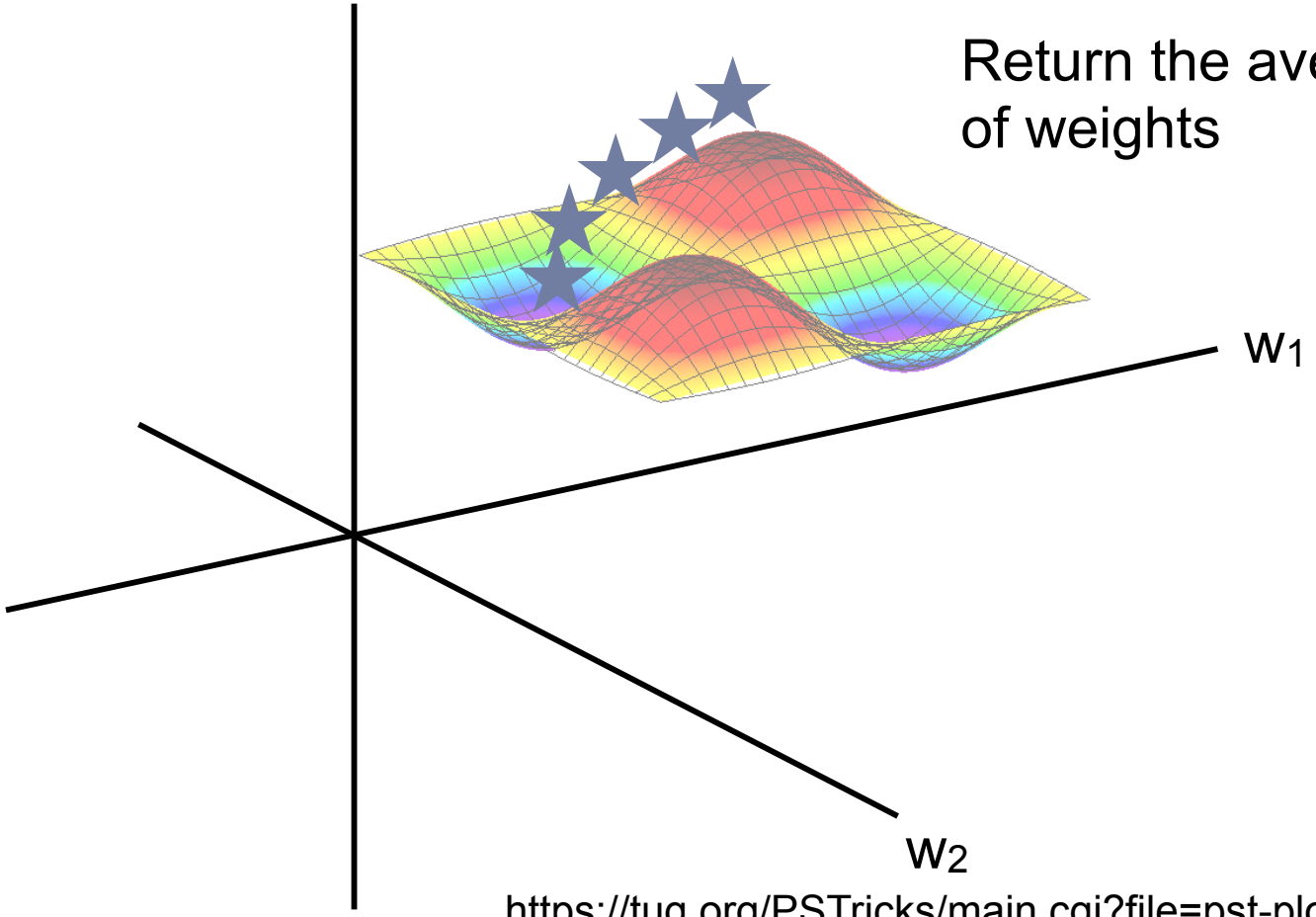
menti.com 7927 6661

Averaged Perceptron

Learning Method

y = How good these weights are (lower is better)

Return the average of weights



https://tug.org/PSTricks/main.cgi?file=pst-plot/3D/contour

COMP 4446 / 5046
Lecture 2, 2025

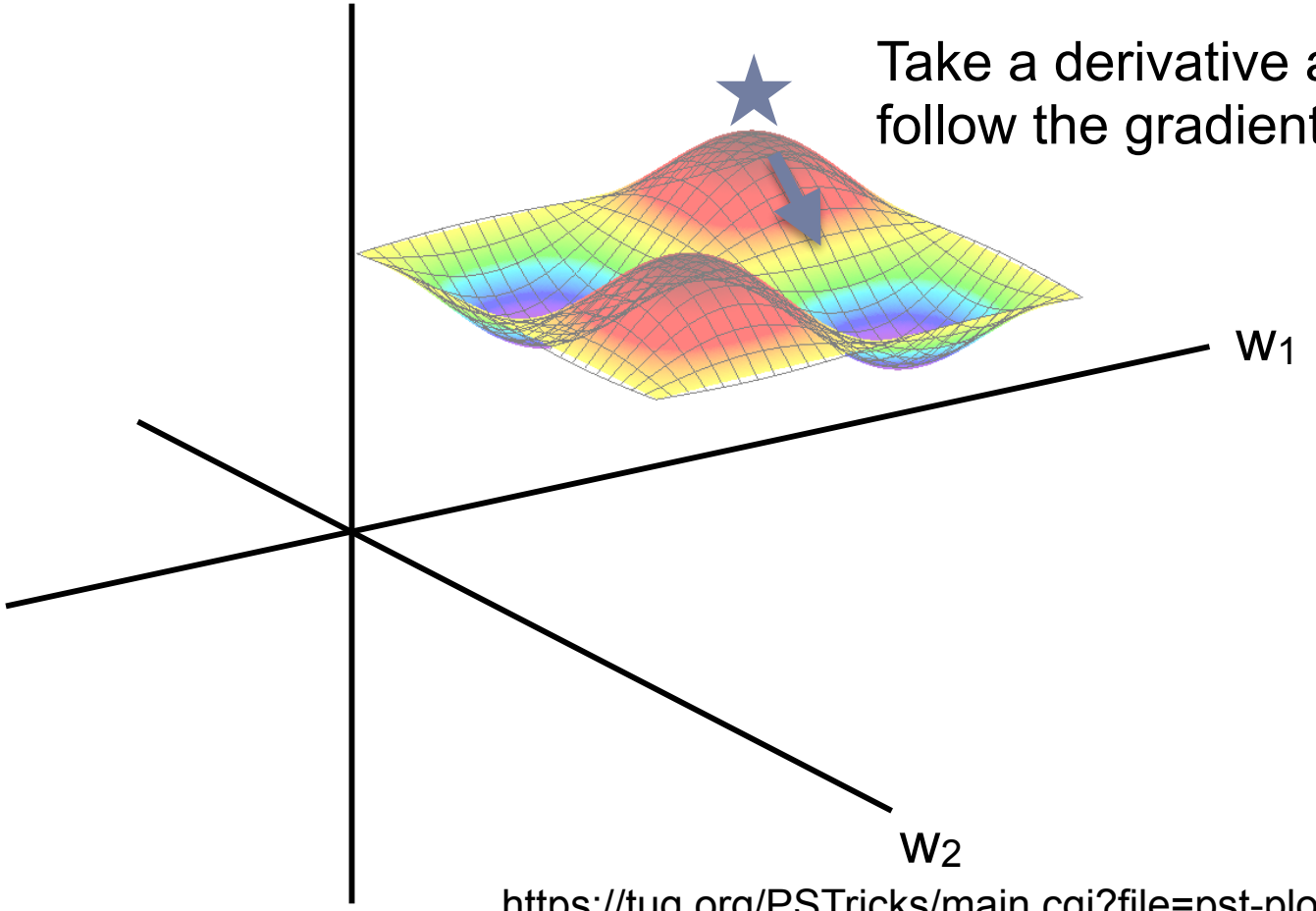
Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Many variations!

Learning Method

y = How good these weights are (lower is better)



https://tug.org/PSTricks/main.cgi?file=pst-plot/3D/contour

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

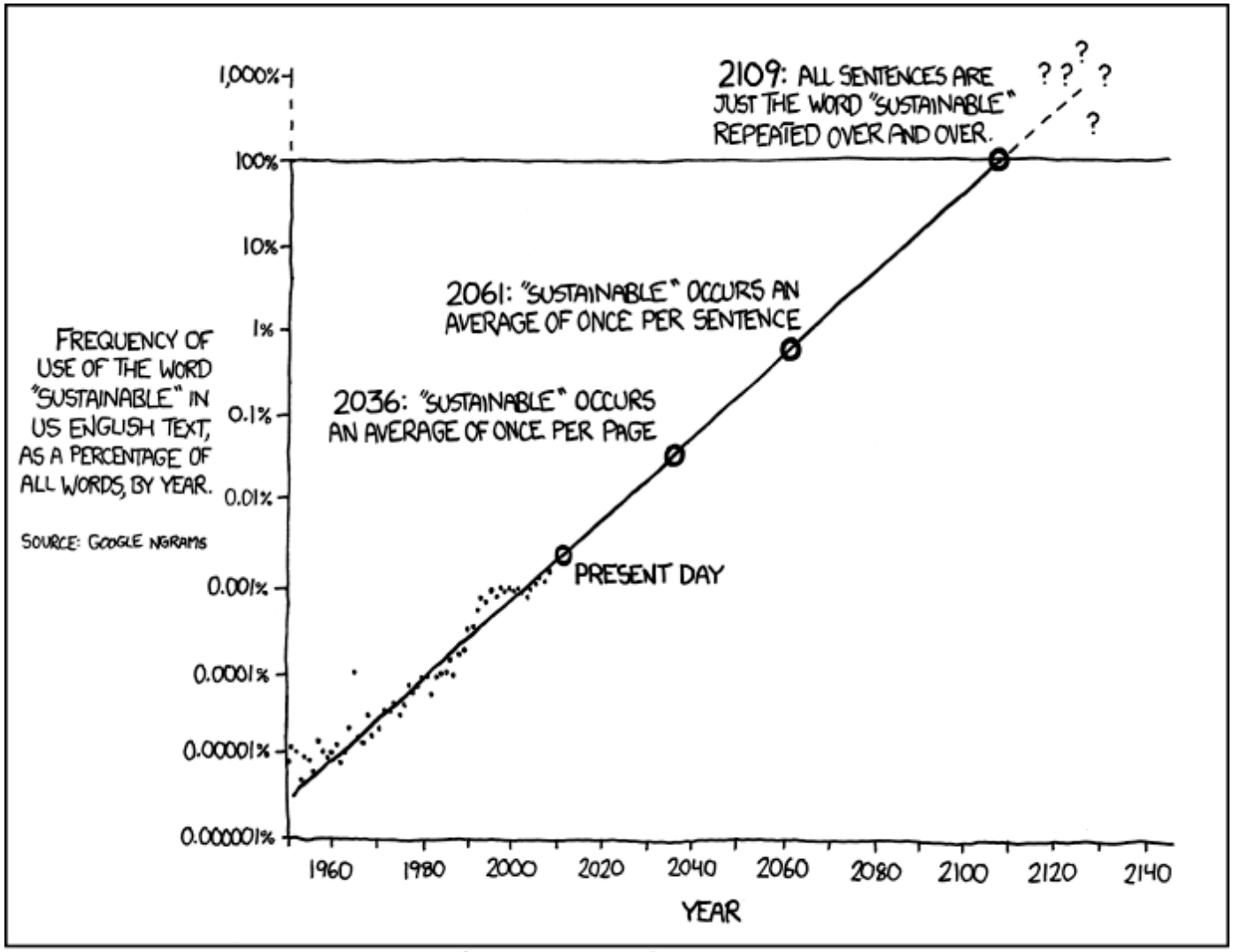
menti.com 7927 6661

NLP systems have five key components

Data	Examples of the language phenomena we want our system to handle
Model	A function that maps (input, output) pairs to scores
Inference Method	A way to make a prediction for an example given a Model
Metric	A function that gives a score to the output produced by a Model given some Data
Learning Method	A way to update a Model given Data a Metric and an Inference Method

3 minute Break - stretch and visit Menti

menti.com
7927 6661



THE WORD "SUSTAINABLE" IS UNSUSTAINABLE.

Sustainable

[Though 100 years is longer than a lot of our resources.]

Source: https://xkcd.com/1007/

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

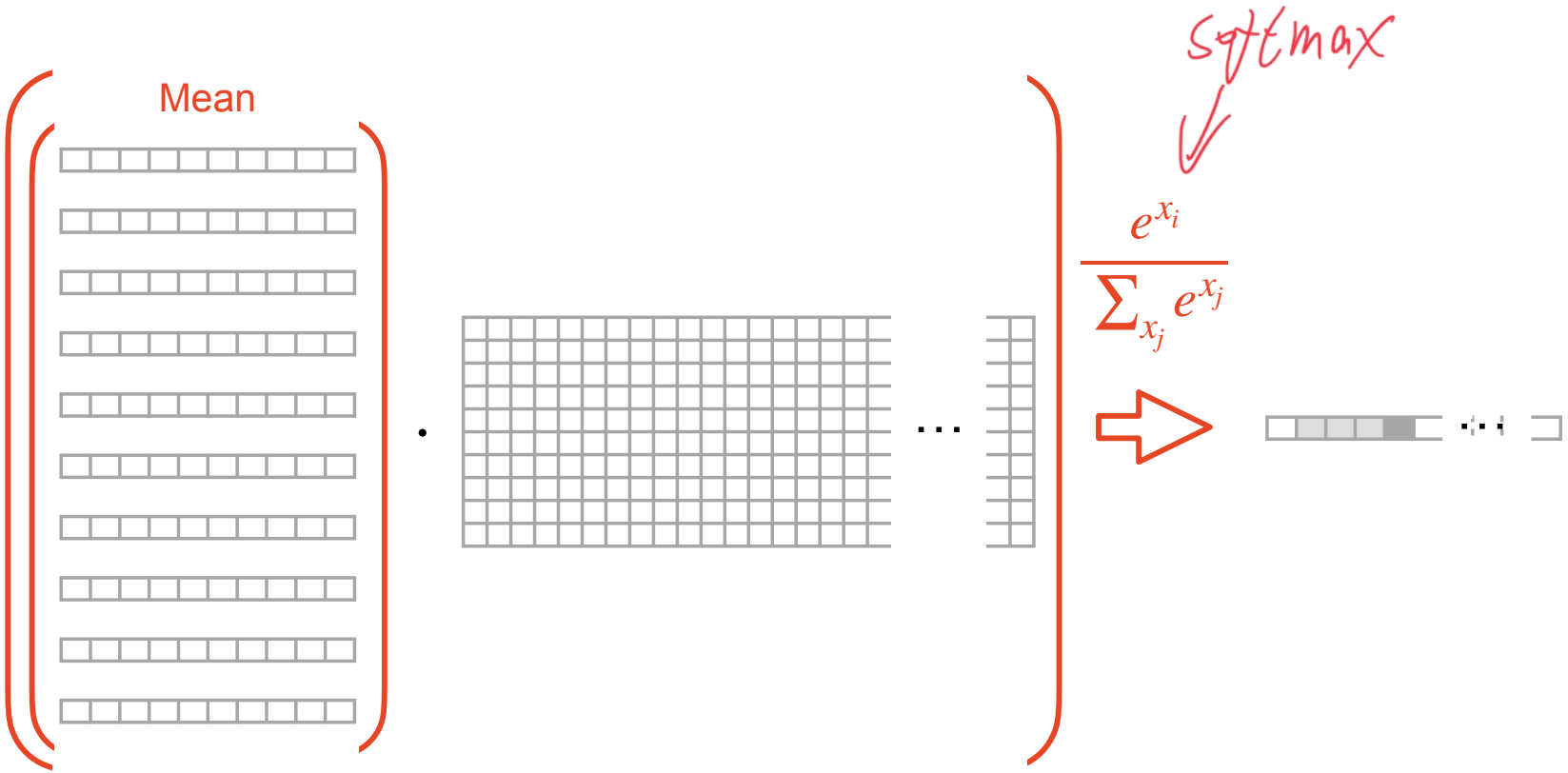
Word Vectors

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Last lecture we described the word2vec algorithm



$$x = \frac{\sum_{\text{words}} \text{word vectors}}{|\text{words}|} \cdot \text{weight matrix}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

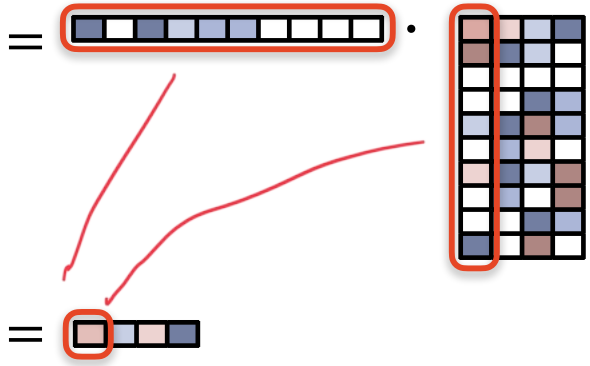
menti.com 7927 6661

We can use those word vectors in the baseline approach

$$\text{scores(doc)} = \text{features(doc)} \cdot \text{weights}$$

Blue = Positive
(darker is more positive)

Red = Negative
(darker is more negative)



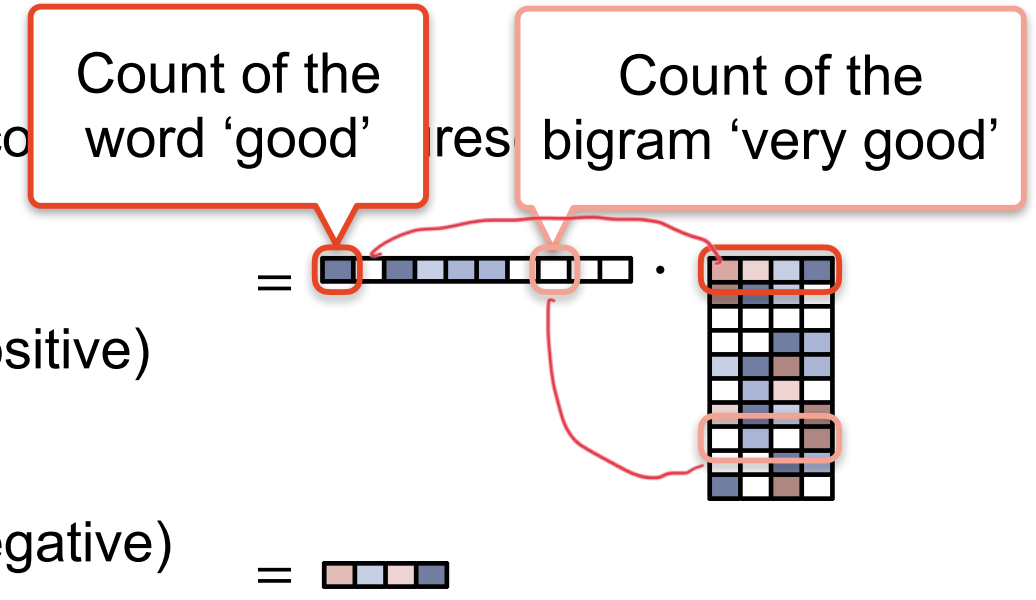
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We can use those word vectors in the baseline approach

sco res



Blue = Positive
(darker is more positive)

Red = Negative
(darker is more negative)

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

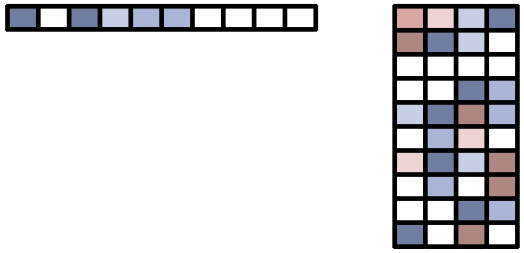
menti.com 7927 6661

We can use those word vectors in the baseline approach

Blue = Positive
(darker is more positive)

Red = Negative
(darker is more negative)

Document
Representation



Task specific mapping
from document
representation to scores

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

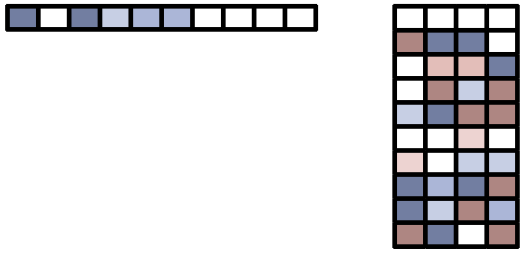
menti.com 7927 6661

We can use those word vectors in the baseline approach

Blue = Positive
(darker is more positive)

Red = Negative
(darker is more negative)

Document
Representation



Task specific mapping
from document
representation to scores

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

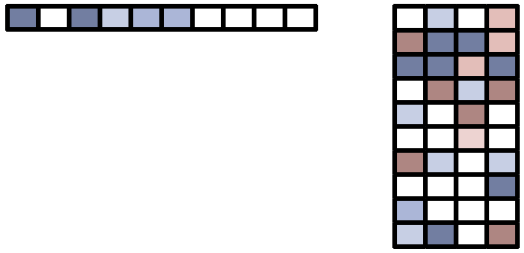
menti.com 7927 6661

We can use those word vectors in the baseline approach

Blue = Positive
(darker is more positive)

Red = Negative
(darker is more negative)

Document
Representation



Task specific mapping
from document
representation to scores

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's consider the task of predicting how cool a word is

Cool	**Not cool**
Bamboozle	The
Suave	Building
Bungalow	Copacetic
Dirigible	Pogo

Task: Given a word, predict which list it should go in.

Does it contain a 'v', 'w', or 'z'?

Is it one of the 100 most frequent words?

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's consider the task of predicting how cool a word is

Does it contain a 'v', 'w', or 'z'?

Is it one of the 100 most frequent words?

Black = True
White = False

Blue = Positive
(darker is more positive)

Red = Negative
(darker is more negative)

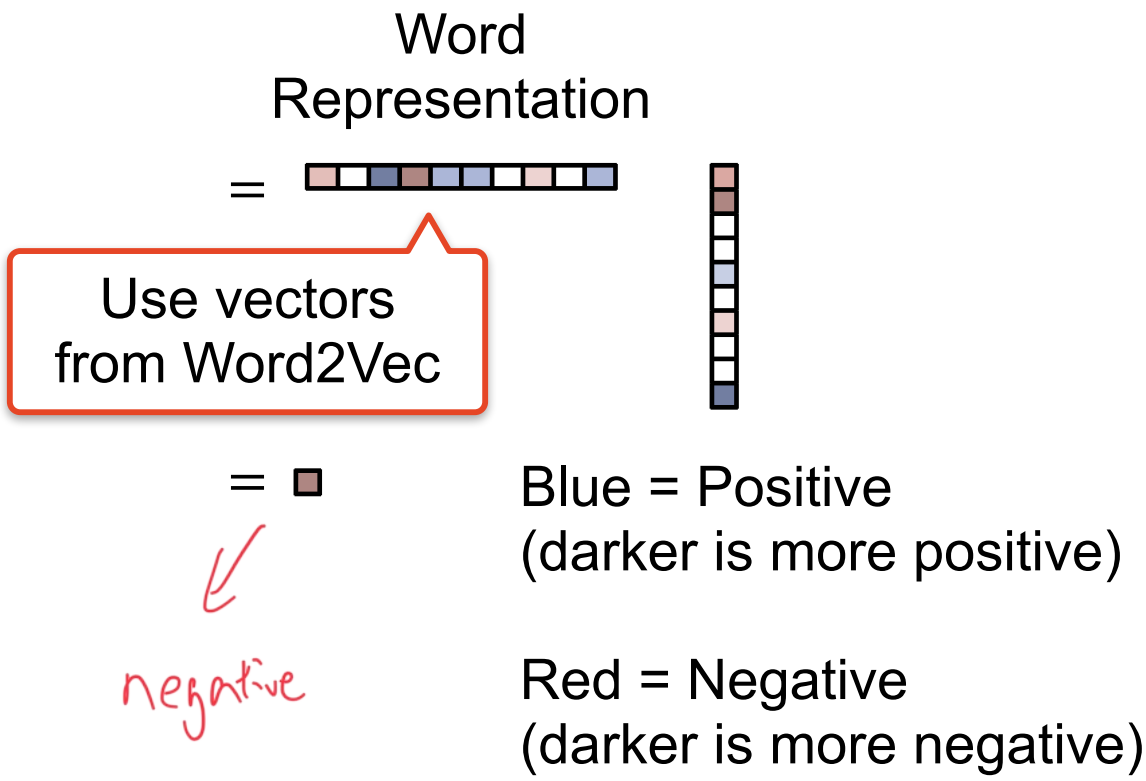
This word representation is:
- Sparse, values are 0 or 1, most are 0
- Human defined, we wrote little rules for each one

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's consider the task of predicting how cool a word is

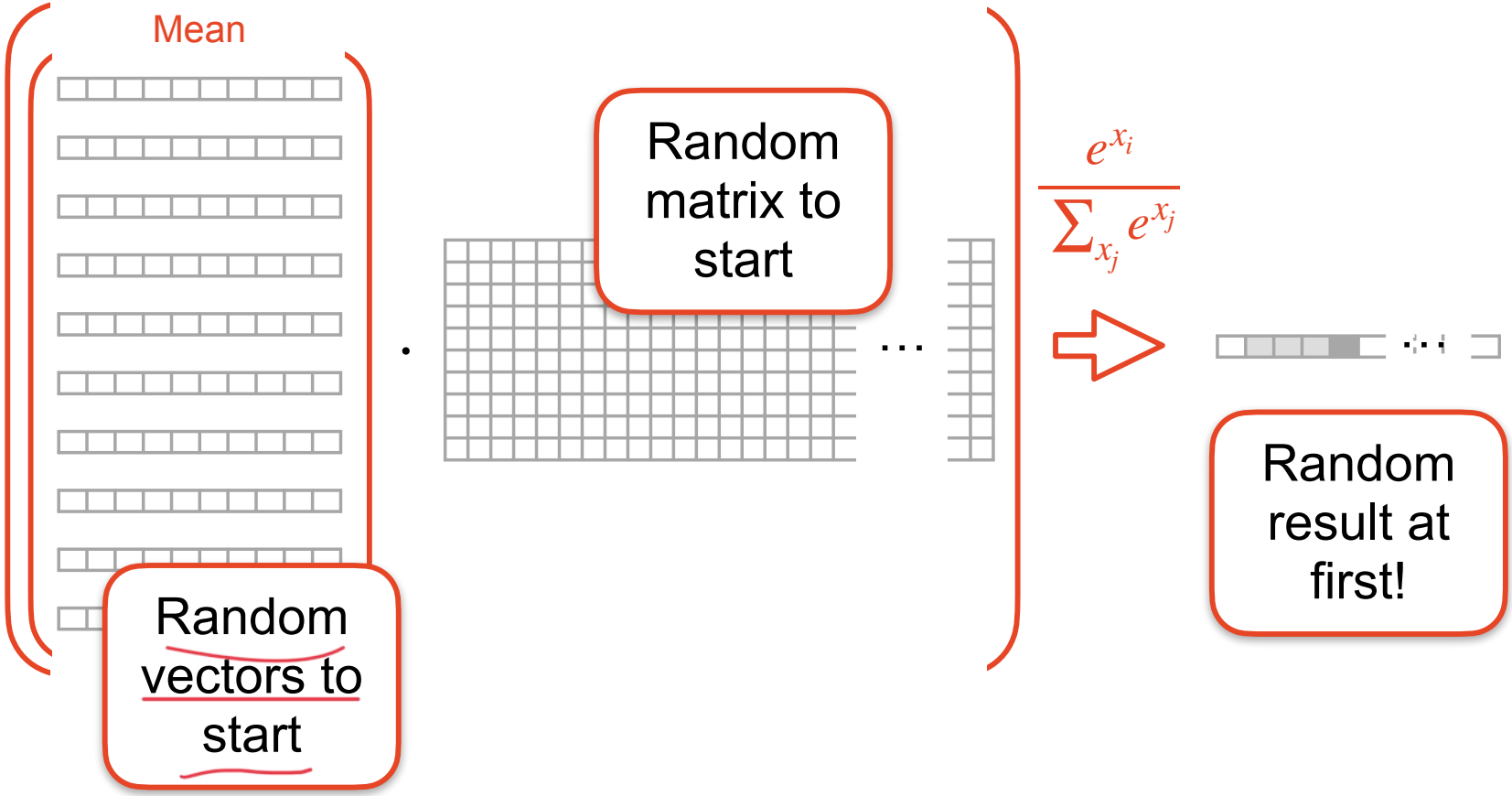


COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We can also now explain learning in word2vec

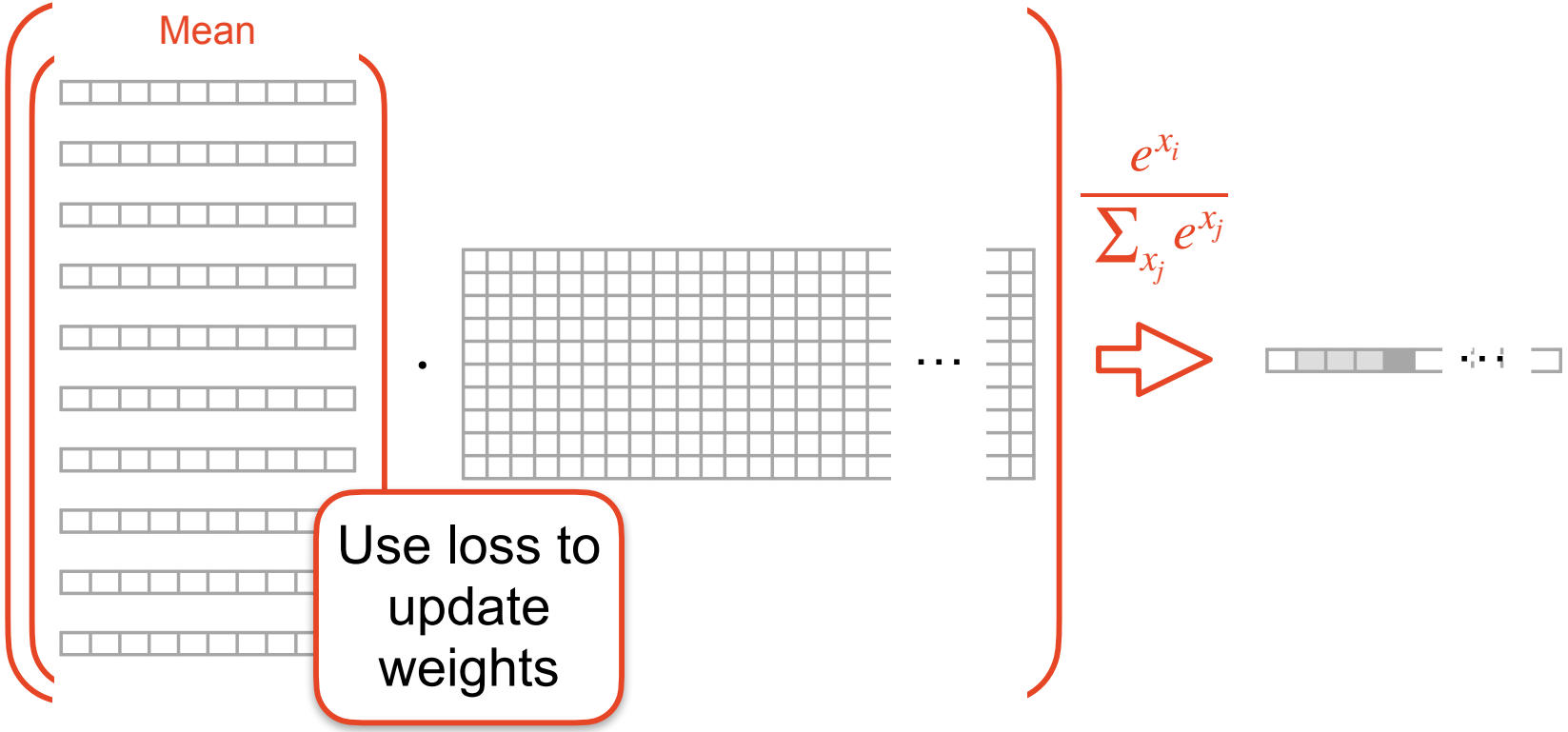


COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We can also now explain learning in word2vec



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

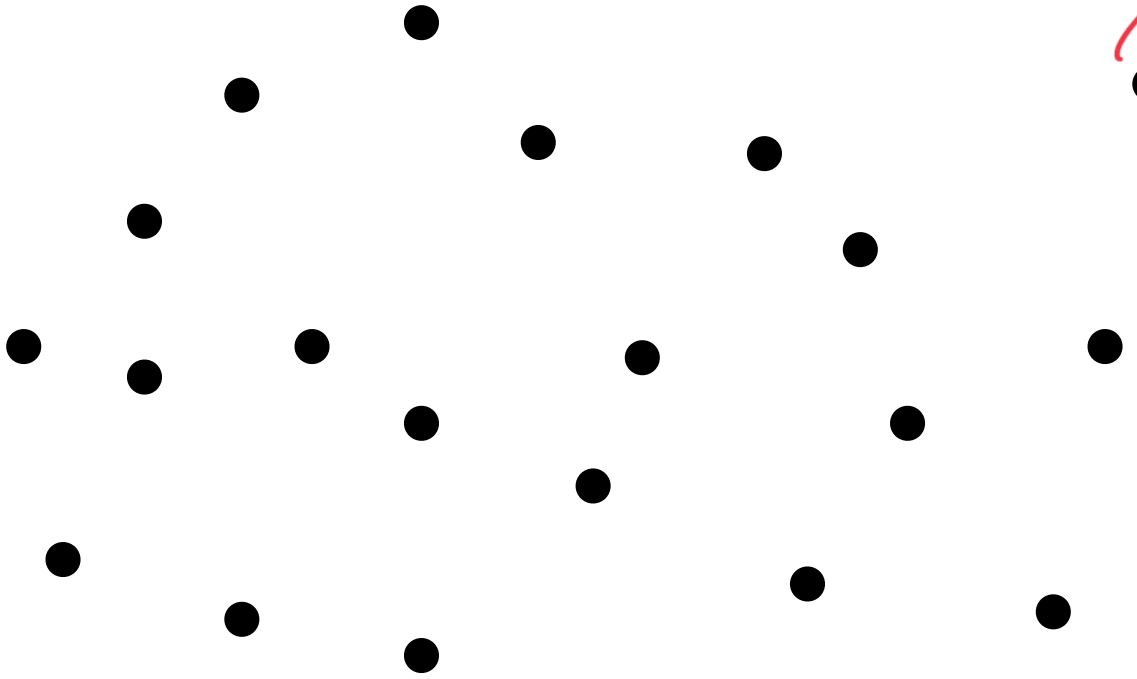
Evaluation

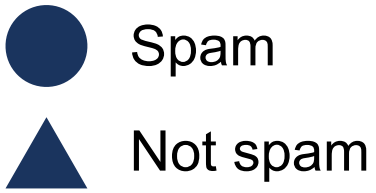
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We group test cases together to calculate metrics



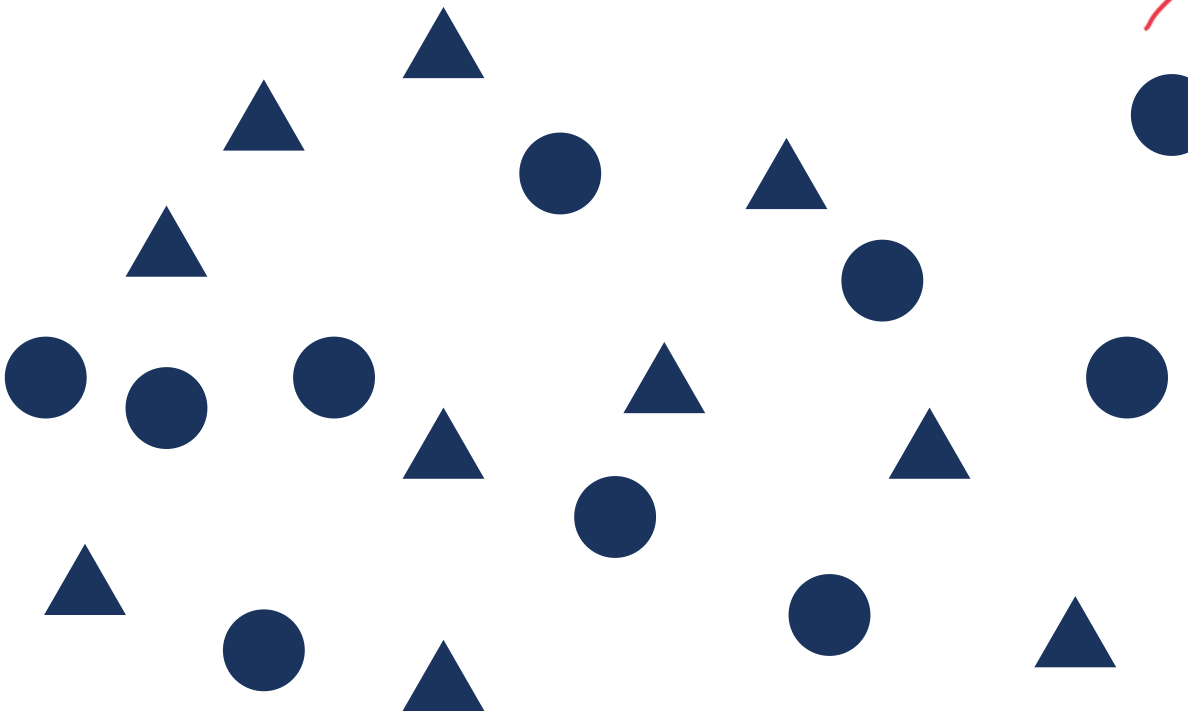


COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We group test cases together to calculate metrics



Spam and predicted Spam

Not spam and predicted Not spam

Spam, predicted Not spam

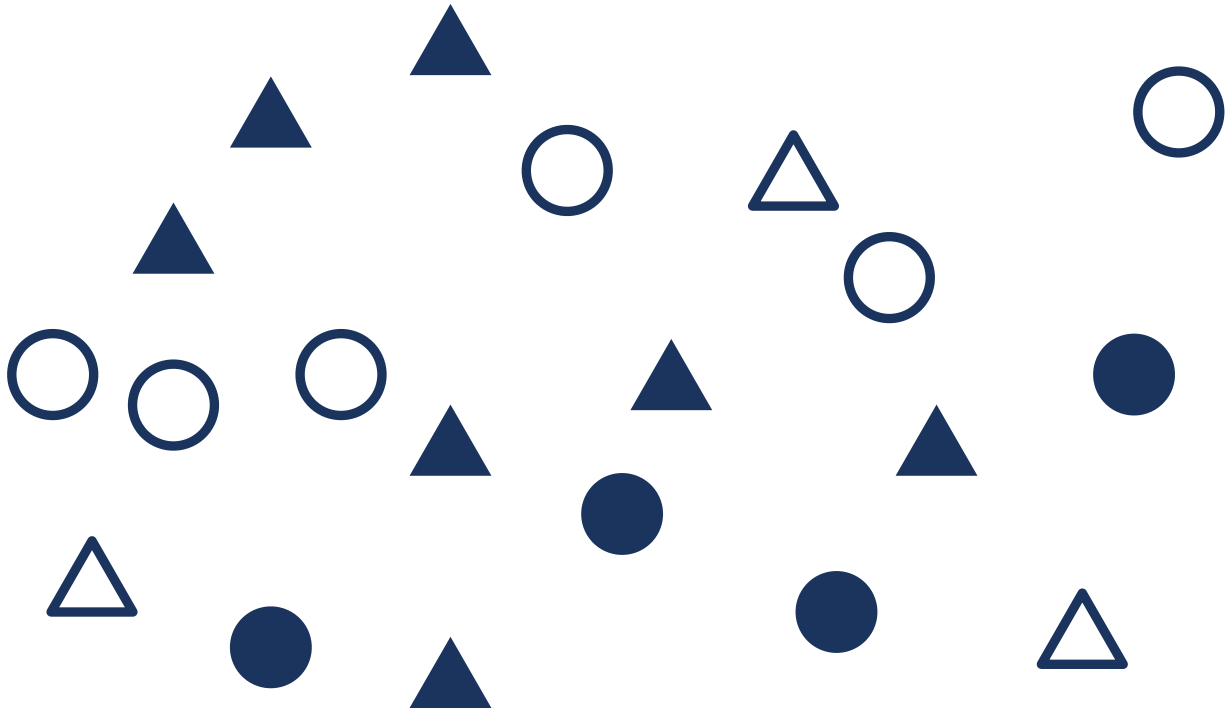
Not spam, predicted Spam

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We group test cases together to calculate metrics



Spam and predicted Spam

Not spam and predicted Not spam

Spam, predicted Not spam

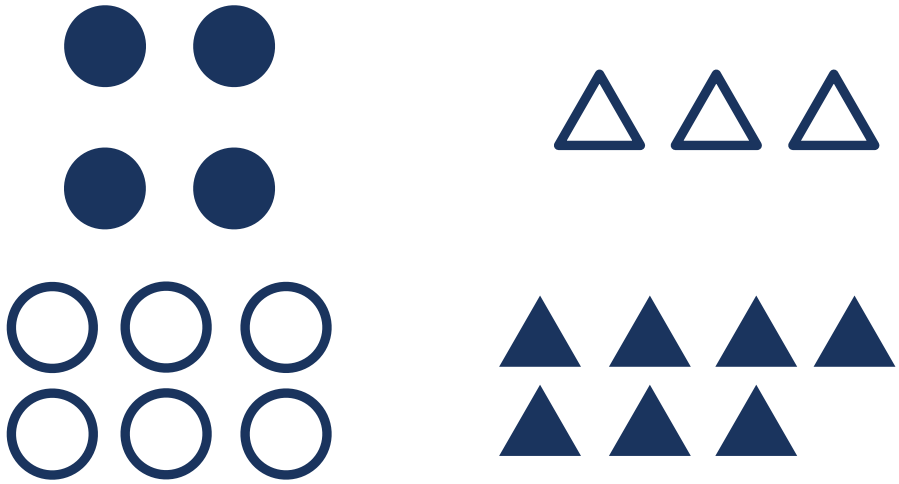
Not spam, predicted Spam

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We group test cases together to calculate metrics



Spam and predicted Spam

Not spam and predicted Not spam

Spam, predicted Not spam

Not spam, predicted Spam

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We group test cases together to calculate metrics

	Answer: Spam	Answer: Not Spam
Guess: Spam	● ● ● ●	△ △ △
Guess: Not Spam	○ ○ ○ ○ ○ ○	▲ ▲ ▲ ▲ ▲ ▲ ▲

● Spam and predicted Spam

▲ Not spam and predicted Not spam

○ Spam, predicted Not spam

△ Not spam, predicted Spam

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We group test cases together to calculate metrics

	Answer: Spam	Answer: Not Spam
Guess: Spam	4 True Positives	△ △ △
Guess: Not Spam	○ ○ ○ ○ ○ ○	▲ ▲ ▲ ▲ ▲ ▲ ▲

● Spam and predicted Spam

▲ Not spam and predicted Not spam

○ Spam, predicted Not spam

△ Not spam, predicted Spam

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We group test cases together to calculate metrics

	Answer: Spam	Answer: Not Spam
Guess: Spam	4 True Positives	3 False Positives
Guess: Not Spam	○ ○ ○ ○ ○ ○	△ △ △ △ △ △ △

● Spam and predicted Spam

▲ Not spam and predicted Not spam

○ Spam, predicted Not spam

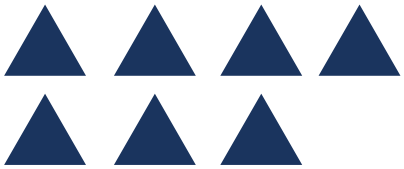
△ Not spam, predicted Spam

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We group test cases together to calculate metrics

	Answer: Spam	Answer: Not Spam
Guess: Spam	4 True Positives	3 False Positives
Guess: Not Spam	6 False Negatives	

 Spam and predicted Spam

 Not spam and predicted Not spam

 Spam, predicted Not spam

 Not spam, predicted Spam

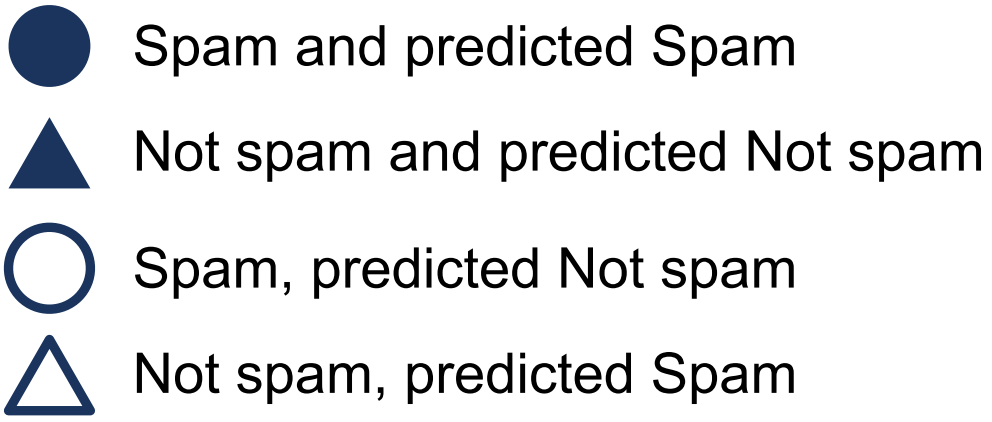
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

We group test cases together to calculate metrics

	Answer: Spam	Answer: Not Spam
Guess: Spam	4 True Positives	3 False Positives
Guess: Not Spam	6 False Negatives	7 True Negatives



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Accuracy is what proportion of cases got the right label

	Answer: Positive	Answer: Negative
Guess: Positive	TP 	FP 
Guess: Negative	FN 	TN 

$$\text{Accuracy} = \frac{\text{Correctly labeled cases}}{\text{All test cases}}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Accuracy is what proportion of cases got the right label

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{Accuracy} = \frac{\text{Correctly labeled cases}}{\text{All test cases}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Quiz?

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Accuracy is what proportion of cases got the right label

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{Accuracy} = \frac{\text{Correctly labeled cases}}{\text{All test cases}}$$

$$= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$= \frac{4 + 7}{4 + 3 + 6 + 7}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Accuracy is what proportion of cases got the right label

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{Accuracy} = \frac{\text{Correctly labeled cases}}{\text{All test cases}}$$

$$= \frac{\text{TP + TN}}{\text{TP + TN + FP + FN}}$$

$$= \frac{4 + 7}{4 + 3 + 6 + 7}$$

$$= \frac{11}{20}$$

$$= 0.55$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Precision is what proportion of cases guessed as positive were actually positive

	Answer: Positive	Answer: Negative
Guess: Positive	TP 	FP 
Guess: Negative	FN 	TN 

$$\text{Precision} = \frac{\text{Cases gussed as positive that are positive}}{\text{Cases guessed as positive}}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Precision is what proportion of cases guessed as positive were actually positive

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{Precision} = \frac{\text{Cases gussed as positive that are positive}}{\text{Cases guessed as positive}}$$

$$= \frac{\text{TP}}{\text{TP + FP}}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Precision is what proportion of cases guessed as positive were actually positive

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{Precision} = \frac{\text{Cases gussed as positive that are positive}}{\text{Cases guessed as positive}}$$

$$= \frac{\text{TP}}{\text{TP + FP}}$$

$$= \frac{4}{4+3}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Precision is what proportion of cases guessed as positive were actually positive

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{Precision} = \frac{\text{Cases gussed as positive that are positive}}{\text{Cases guessed as positive}}$$

$$= \frac{\text{TP}}{\text{TP + FP}}$$

$$= \frac{4}{4+3}$$

$$= \frac{4}{7}$$

$$= 0.57$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Recall is what proportion of cases that should be positive were guessed as positive

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{Recall} = \frac{\text{Cases gussed as positive that are positive}}{\text{Cases that are positive}}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Recall is what proportion of cases that should be positive were guessed as positive

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{Recall} = \frac{\text{Cases gussed as positive that are positive}}{\text{Cases that are positive}}$$

$$= \frac{\text{TP}}{\text{TP + FN}}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Recall is what proportion of cases that should be positive were guessed as positive

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{Recall} = \frac{\text{Cases gussed as positive that are positive}}{\text{Cases that are positive}}$$

$$= \frac{\text{TP}}{\text{TP + FN}}$$

$$= \frac{4}{4+6}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Recall is what proportion of cases that should be positive were guessed as positive

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{Recall} = \frac{\text{Cases gussed as positive that are positive}}{\text{Cases that are positive}}$$

$$= \frac{\text{TP}}{\text{TP + FN}}$$

$$= \frac{4}{4+6}$$

$$= \frac{4}{10}$$

$$= 0.40$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

F-Score is the harmonic mean of Precision and Recall

mean

$$\text{F-Score} = \frac{2}{\frac{1}{P} + \frac{1}{R}}$$

mean

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

F-Score is the harmonic mean of Precision and Recall

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{F-Score} = \frac{2}{\frac{1}{P} + \frac{1}{R}}$$

$$= \frac{2PR}{P + R}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

F-Score is the harmonic mean of Precision and Recall

	Answer: Positive	Answer: Negative
Guess: Positive	TP 	FP 
Guess: Negative	FN 	TN 

$$\text{F-Score} = \frac{2}{\frac{1}{P} + \frac{1}{R}}$$

$$= \frac{2PR}{P + R}$$

$$= \frac{2TP}{2TP + FP + FN}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

F-Score is the harmonic mean of Precision and Recall

	Answer: Positive	Answer: Negative
Guess: Positive	TP ●● ●●	FP △ △ △
Guess: Negative	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

$$\text{F-Score} = \frac{2}{\frac{1}{P} + \frac{1}{R}}$$

$$= \frac{2PR}{P+R} \qquad = \frac{2 * 0.4 * 0.57}{0.4 + 0.57} = 0.47$$

$$= \frac{2TP}{2TP + FP + FN} \qquad = \frac{2 * 4}{2 * 4 + 3 + 6} = \frac{8}{17} = 0.47$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Generalising to more than one class

		True Answer	
		Yes	No
Guess	Yes	TP ●● ●●	FP △ △ △
	No	FN ○○○ ○○○	TN ▲▲▲▲ ▲▲▲

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Generalising to more than one class

		True Answer	
		Yes	No
Guess	Yes	TP	FP
	No	FN	TN

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Generalising to more than one class

		True Answer				
		A	B	C	…	None
Guess	A	TP				FP
	B		TP			FP
	C			TP		FP
	…				TP	FP
	None	FN	FN	FN	FN	TN

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Generalising to more than one class

		True Answer			
		A	B	C	…
Guess	A	TP			
	B		TP		
	C			TP	
	…				TP

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Generalising to more than one class

		True Answer			
		A	B	C	…
Guess	A	TP			
	B		TP		
	C			TP	
	…				TP

$$\text{Accuracy} = \frac{\text{Correctly labeled cases}}{\text{All test cases}}$$

$$= \frac{\text{TP + TN}}{\text{All test cases}}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Generalising to more than one class

		True Answer			
		A	B	C	…
Guess	A	TP			
	B		TP		
	C			TP	
	…				TP

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's define P and R for one class

		True Answer			
		A	B	C	…
Guess	A	TP			
	B		TP_B		
	C			TP	
	…				TP

$$\text{Precision} = \frac{\text{Cases gussed as positive that are positive}}{\text{Cases guessed as positive}}$$

$$= \frac{TP}{TP + FP}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's define P and R for one class

		True Answer			
		A	B	C	…
Guess	A	TP			
	B		TP_B		
	C			TP	
	…				TP

$$\text{Precision for B} = \frac{\text{Cases gussed as B that are B}}{\text{Cases guessed as B}}$$

$$= \frac{TP_B}{TP_B + FP_B}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's define P and R for one class

		True Answer			
		A	B	C	…
Guess	A	TP			
	B	FP_B	TP_B	FP_B	FP_B
	C			TP	
	…				TP

$$\text{Precision for B} = \frac{\text{Cases gussed as B that are B}}{\text{Cases guessed as B}}$$

$$= \frac{TP_B}{TP_B + FP_B}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's define P and R for one class

		True Answer			
		A	B	C	…
Guess	A	TP			
	B	FP_B	TP_B	FP_B	FP_B
	C			TP	
	…				TP

$$\text{Recall for B} = \frac{\text{Cases gussed as B that are B}}{\text{Cases that are B}}$$

$$= \frac{TP_B}{TP_B + FN_B}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's define P and R for one class

		True Answer			
		A	B	C	…
Guess	A	TP	FN_B		
	B	FP_B	TP_B	FP_B	FP_B
	C		FN_B	TP	
	…		FN_B		TP

$$\text{Recall for B} = \frac{\text{Cases gussed as B that are B}}{\text{Cases that are B}}$$

$$= \frac{TP_B}{TP_B + FN_B}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's define P and R for one class

		True Answer			
		A	B	C	…
Guess	A	TP	FN_B		
	B	FP_B	TP_B	FP_B	FP_B
	C		FN_B	TP	
	…		FN_B		TP

$$\text{Recall for C} = \frac{\text{Cases gussed as C that are C}}{\text{Cases that are C}}$$

$$= \frac{TP_C}{TP_C + FN_C}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Let's define P and R for one class

		True Answer			
		A	B	C	…
Guess	A	TP	FN_B	FN_C	
	B	FP_B	TP_B	FN_C	FP_B
	C	FP_C	FP_C	TP	FP_C
	…		FN_B	FN_C	TP

$$\text{Recall for C} = \frac{\text{Cases gussed as C that are C}}{\text{Cases that are C}}$$

$$= \frac{TP_C}{TP_C + FN_C}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

How do we combine these scores together?

$$\text{Precision for A} = \frac{TP_A}{TP_A + FP_A} \qquad \text{Recall for A} = \frac{TP_A}{TP_A + FN_A}$$

$$\text{Precision for B} = \frac{TP_B}{TP_B + FP_B} \qquad \text{Recall for B} = \frac{TP_B}{TP_B + FN_B}$$

$$\text{Precision for C} = \frac{TP_C}{TP_C + FP_C} \qquad \text{Recall for C} = \frac{TP_C}{TP_C + FN_C}$$

…

$$\text{Precision}_{micro} = \frac{TP_A + TP_B + TP_C + ...}{TP_A + FP_A + TP_B + FP_B + TP_C + FP_C + ...}$$

$$\text{Recall}_{micro} = \frac{TP_A + TP_B + TP_C + ...}{TP_A + FN_A + TP_B + FN_B + TP_C + FN_C + ...}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

How do we combine these scores together?

$$\text{Precision}_{\text{macro}} = \frac{\frac{TP_A}{TP_A + FP_A} + \frac{TP_B}{TP_B + FP_B} + \frac{TP_C}{TP_C + FP_C} + \ldots}{\text{Number of classes}}$$

$$\text{Recall}_{\text{macro}} = \frac{\frac{TP_A}{TP_A + FN_A} + \frac{TP_B}{TP_B + FN_B} + \frac{TP_C}{TP_C + FN_C} + \ldots}{\text{Number of classes}}$$

$$\text{Precision}_{\text{micro}} = \frac{TP_A + TP_B + TP_C + \ldots}{TP_A + FP_A + TP_B + FP_B + TP_C + FP_C + \ldots}$$

$$\text{Recall}_{\text{micro}} = \frac{TP_A + TP_B + TP_C + \ldots}{TP_A + FN_A + TP_B + FN_B + TP_C + FN_C + \ldots}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

How do we combine these scores together?

$$\text{Precision}_{\text{macro}} = \frac{\sum_{l \in \text{labels}} \frac{\text{TP}_l}{\text{TP}_l + \text{FP}_l}}{\text{Number of classes}} = \frac{\sum_{l \in \text{labels}} \text{Precision}_l}{\text{Number of classes}}$$

$$\text{Recall}_{\text{macro}} = \frac{\sum_{l \in \text{labels}} \frac{\text{TP}_l}{\text{TP}_l + \text{FN}_l}}{\text{Number of classes}} = \frac{\sum_{l \in \text{labels}} \text{Recall}_l}{\text{Number of classes}}$$

$$\text{Precision}_{\text{micro}} = \frac{\text{TP}_A + \text{TP}_B + \text{TP}_C + ...}{\text{TP}_A + \text{FP}_A + \text{TP}_B + \text{FP}_B + \text{TP}_C + \text{FP}_C + ...}$$

$$\text{Recall}_{\text{micro}} = \frac{\text{TP}_A + \text{TP}_B + \text{TP}_C + ...}{\text{TP}_A + \text{FN}_A + \text{TP}_B + \text{FN}_B + \text{TP}_C + \text{FN}_C + ...}$$

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

How do we combine these scores together?

$$\text{Precision}_{\text{macro}} = \frac{\sum_{l \in \text{labels}} \frac{\text{TP}_l}{\text{TP}_l + \text{FP}_l}}{\text{Number of classes}} = \frac{\sum_{l \in \text{labels}} \text{Precision}_l}{\text{Number of classes}}$$

$$\text{Recall}_{\text{macro}} = \frac{\sum_{l \in \text{labels}} \frac{\text{TP}_l}{\text{TP}_l + \text{FN}_l}}{\text{Number of classes}} = \frac{\sum_{l \in \text{labels}} \text{Recall}_l}{\text{Number of classes}}$$

$$\text{Precision}_{\text{micro}} = \frac{\sum_{l \in \text{labels}} \text{TP}_l}{\sum_{l \in \text{labels}} \text{TP}_l + \text{FP}_l}$$

$$\text{Recall}_{\text{micro}} = \frac{\sum_{l \in \text{labels}} \text{TP}_l}{\sum_{l \in \text{labels}} \text{TP}_l + \text{FN}_l}$$

menti.com 7927 6661

How do we combine these scores together?

2 者的区别?

Less common

$$\text{Precision}_{\text{macro}} = \frac{\sum_{l \in \text{labels}} \frac{\text{TP}_l}{\text{TP}_l + \text{FP}_l}}{\text{Number of classes}} = \frac{\sum_{l \in \text{labels}} \text{Precision}_l}{\text{Number of classes}}$$

$$\text{Recall}_{\text{macro}} = \frac{\sum_{l \in \text{labels}} \frac{\text{TP}_l}{\text{TP}_l + \text{FN}_l}}{\text{Number of classes}} = \frac{\sum_{l \in \text{labels}} \text{Recall}_l}{\text{Number of classes}}$$

More common

$$\text{Precision}_{\text{micro}} = \frac{\sum_{l \in \text{labels}} \text{TP}_l}{\sum_{l \in \text{labels}} \text{TP}_l + \text{FP}_l}$$

$$\text{Recall}_{\text{micro}} = \frac{\sum_{l \in \text{labels}} \text{TP}_l}{\sum_{l \in \text{labels}} \text{TP}_l + \text{FN}_l}$$

badly on 1 label

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

A multi-class example

		True Answer		
		Urgent Email	Normal Email	Spam
Guess	Urgent Email	8	10	1
	Normal Email	5	60	50
	Spam	3	30	200

$$P = \frac{8}{8+10+1} \qquad P = \frac{60}{5+60+50} \qquad P = \frac{200}{3+30+200}$$

$$R = \frac{8}{8+5+3} \qquad R = \frac{60}{10+60+30} \qquad R = \frac{200}{1+50+200}$$

$$P_{micro} = \frac{8+60+200}{(8+10+1)+(5+60+50)+(3+30+200)} = 0.73$$

$$R_{micro} = \frac{8+60+200}{(8+5+3)+(10+60+30)+(1+50+200)} = 0.73$$

These are always equal when there is no 'None' class.

Lecture Slides at Stanford

for micro

Pmacro ≠ Rmacro

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Which metric do you care about most?

对每个 task 都不同

Spam Detection
False-Positives (miss a real message) are worse than False-Negatives (have some more spam to delete)

Providing Document in Court
False-Negative (leaving out a document you are meant to share) is worse than False-Positive (sharing documents that are irrelevant)

If you flip the task then what matters flips
e.g., ‘spam detection’ could be ‘relevant detection’, now recall matters more than precision

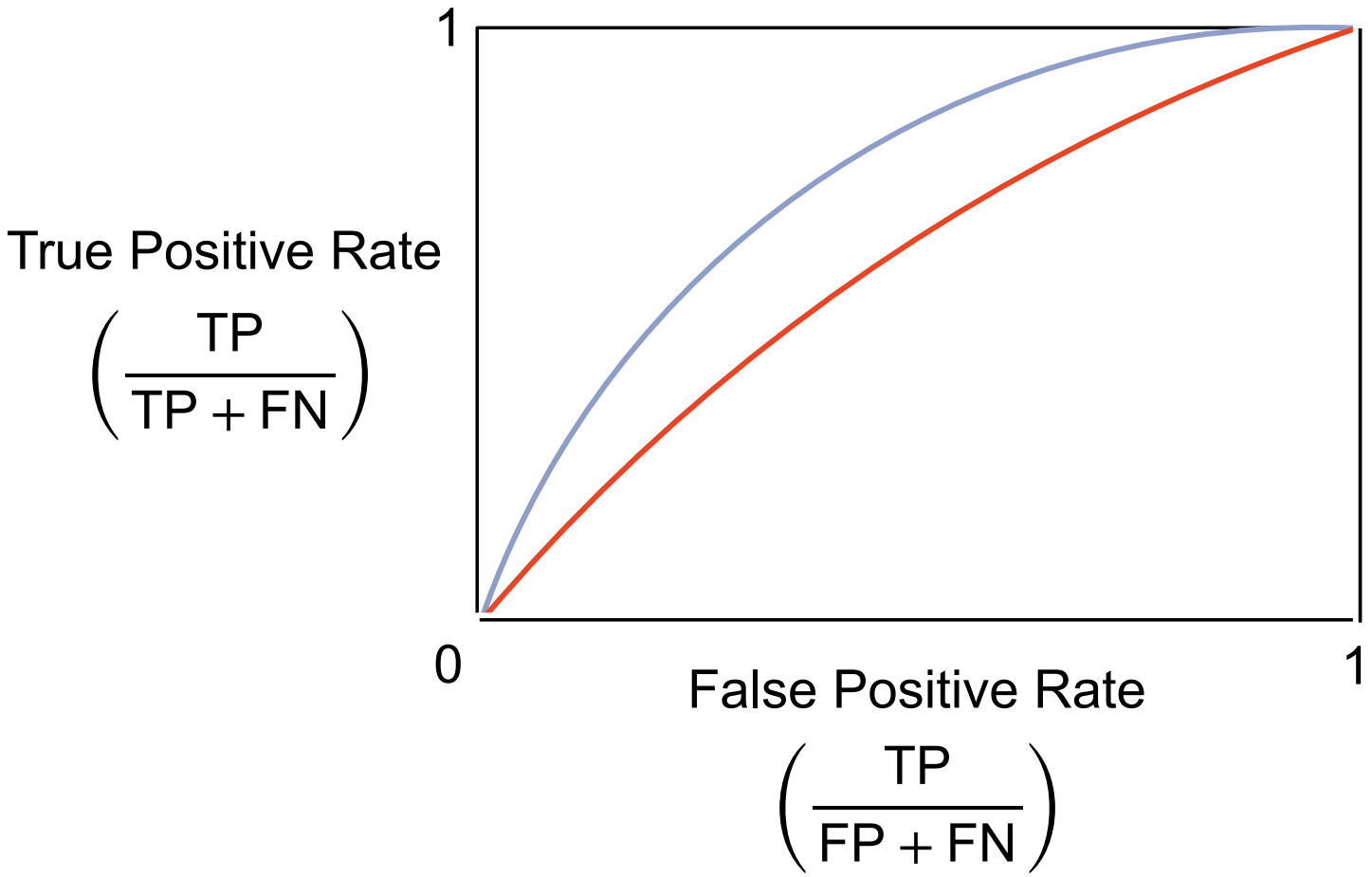
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Sometimes we can adjust tradeoffs

ROC: Receiver operating characteristic curve



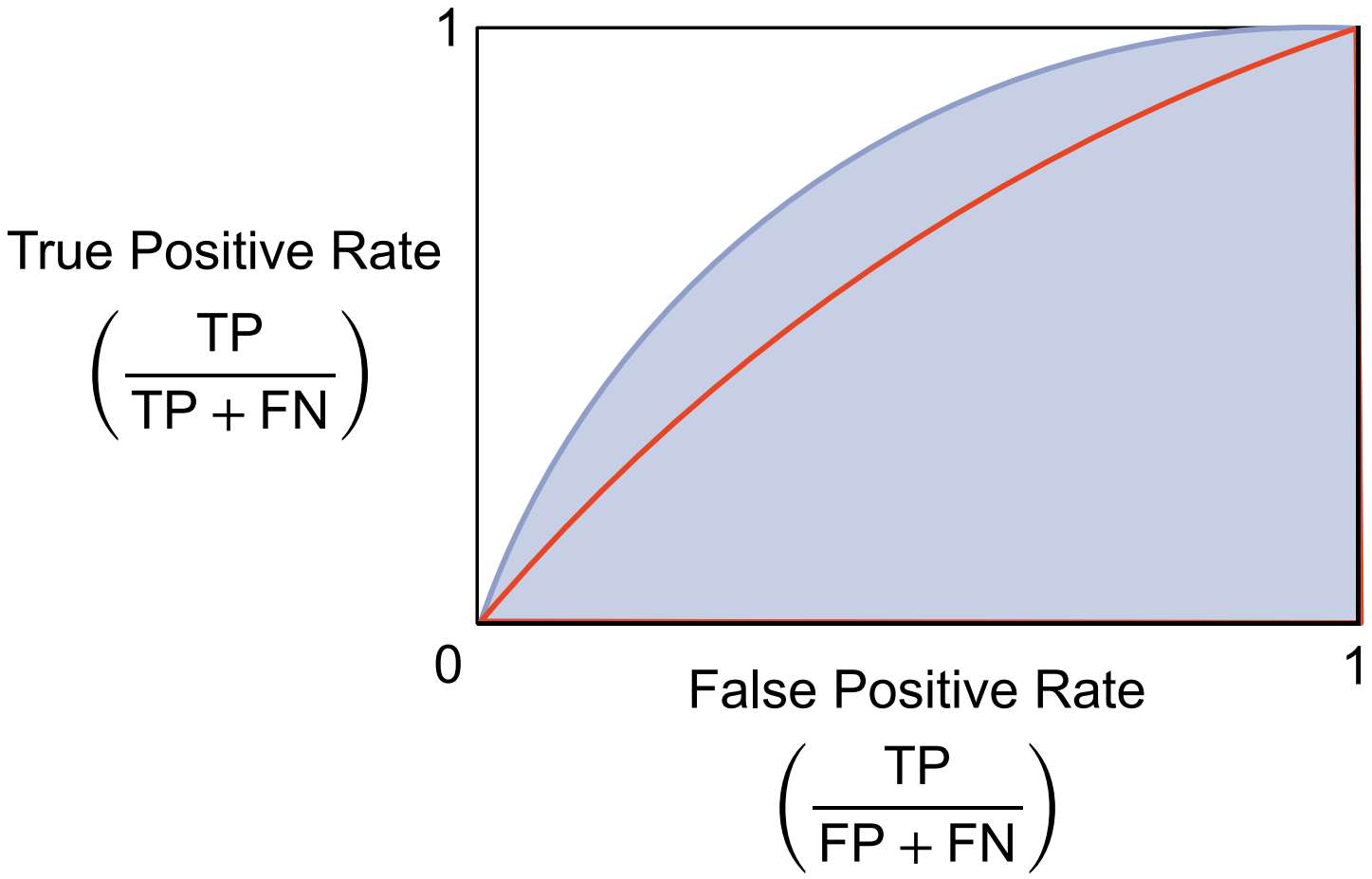
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Sometimes we can adjust tradeoffs

ROC: Receiver operating characteristic curve



COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

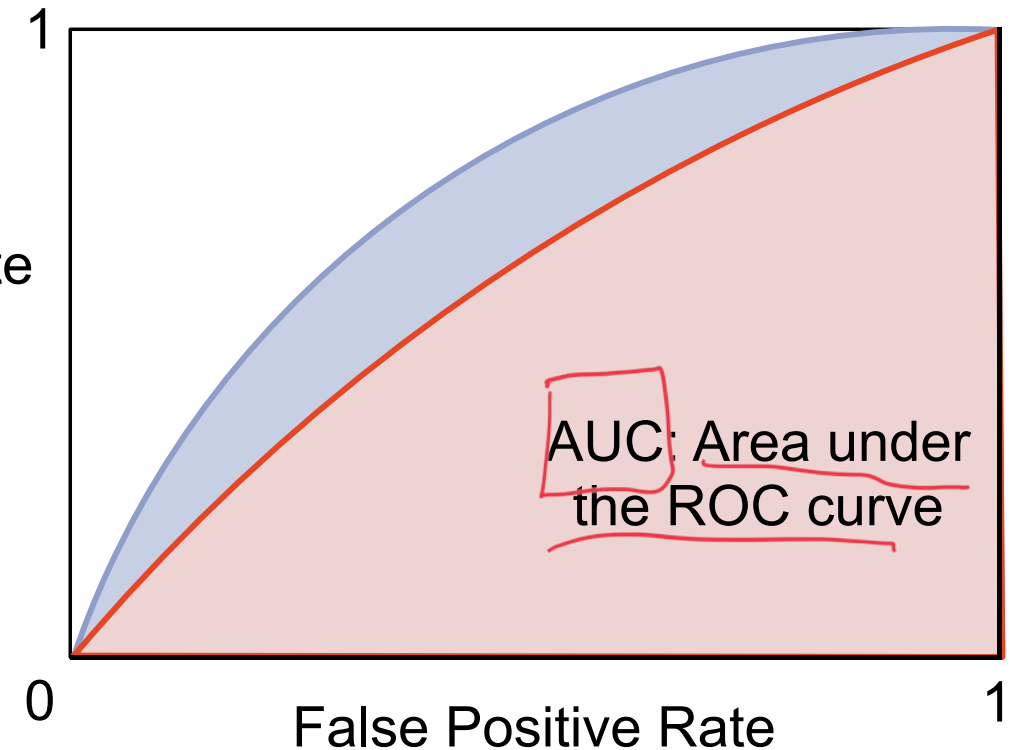
menti.com 7927 6661

Sometimes we can adjust tradeoffs

ROC: Receiver operating characteristic curve

True Positive Rate

$$\left(\frac{TP}{TP + FN}\right)$$



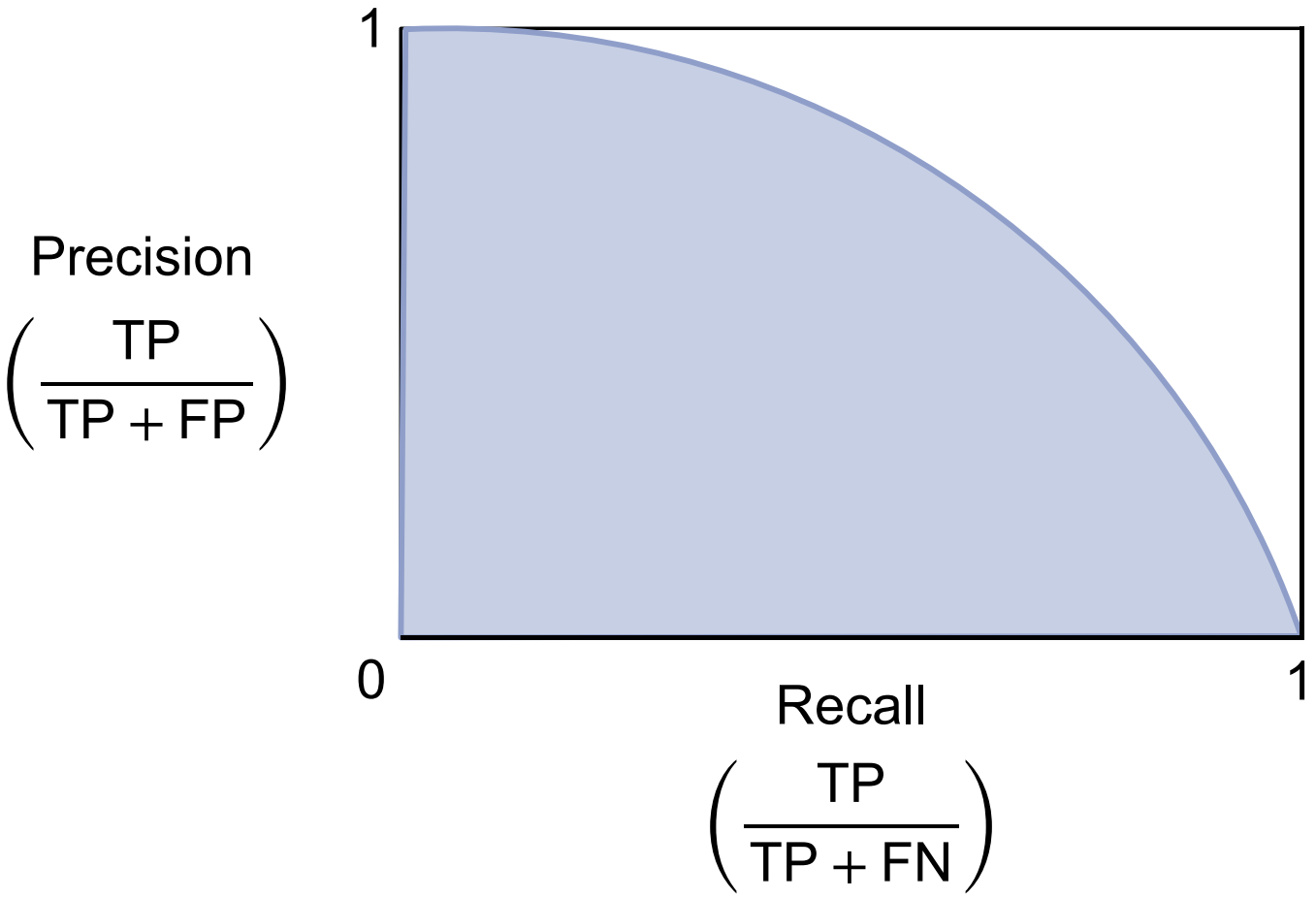
COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Sometimes we can adjust tradeoffs

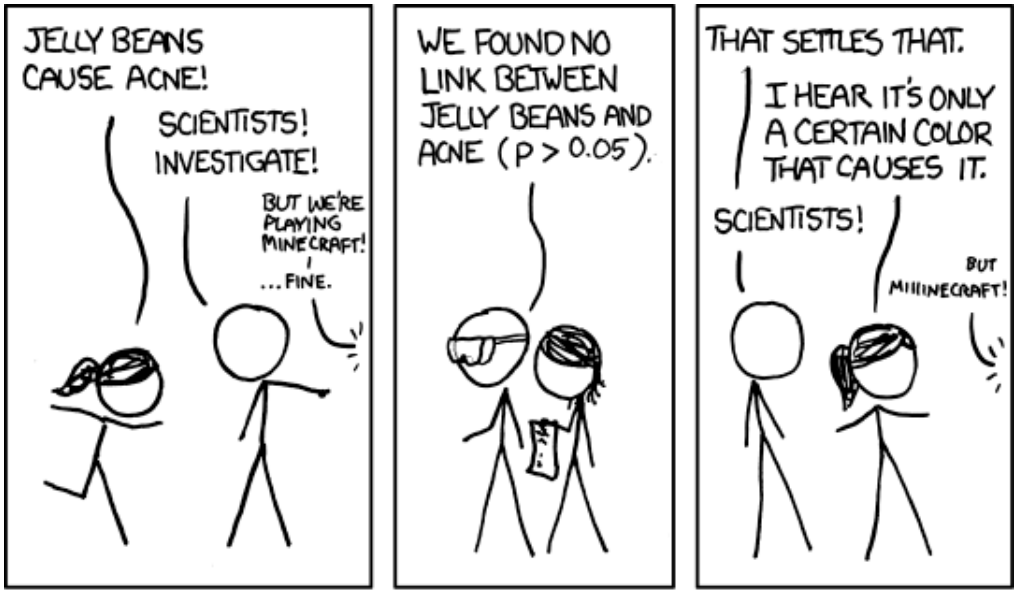
PRC: Precision Recall Curve

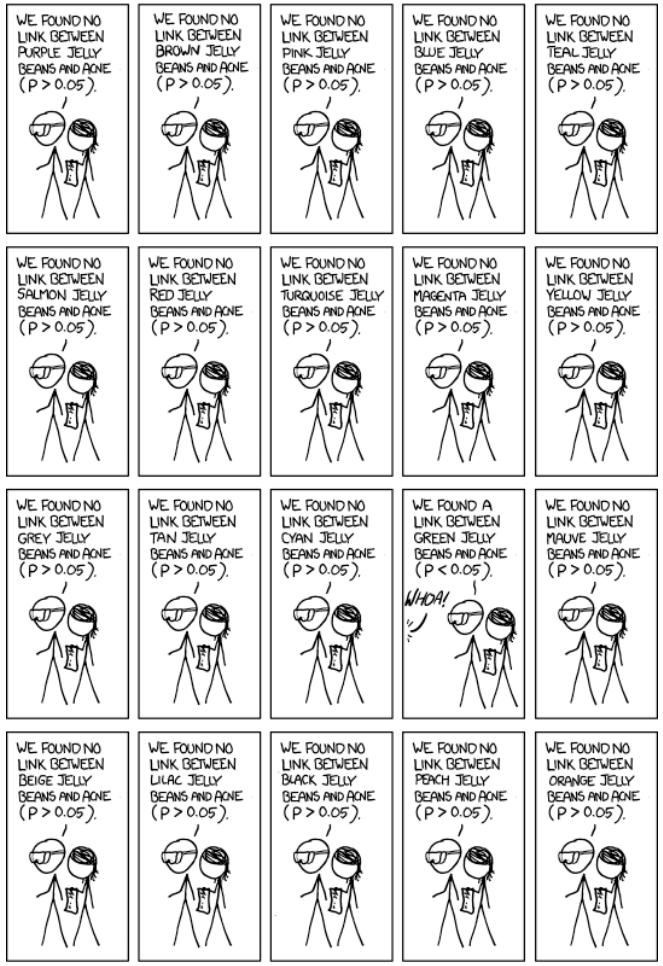


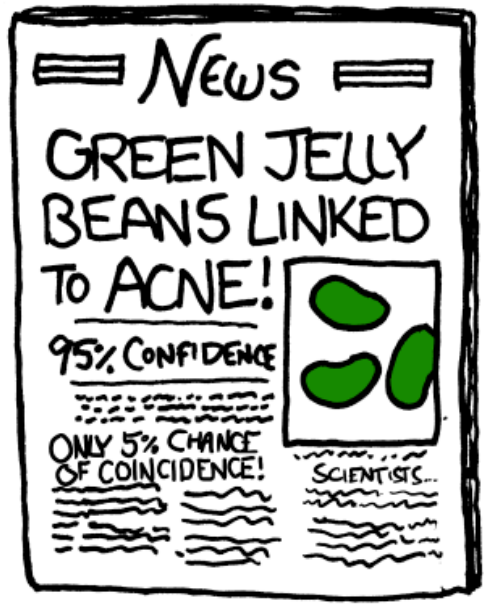
3 minute Break - stretch and visit Menti

menti.com
7927 6661

Significant







["So, uh, we did the green study again and got no link. It was probably a—" "RESEARCH CONFLICTED ON GREEN JELLY BEAN ACNE LINK; MORE STUDY RECOMMENDED!"]

Source: https://xkcd.com/882/

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

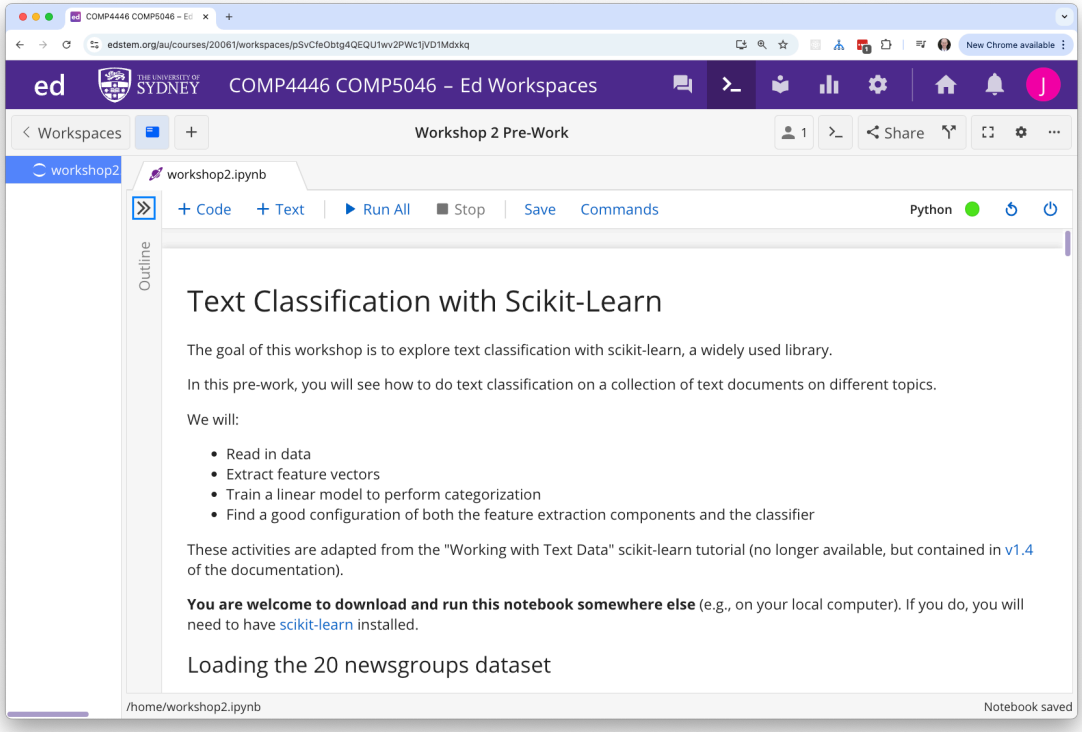
Workshop Preview

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Pre-Workshop:



In Workshop:

Develop a classifier
to detect deception
in Diplomacy

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Muddy Card

Bad examples (still got credit in week 1, not so generous in future):

everything….

Real good lecture

maybe the room is so hot

Good examples:

How does dimensionality reduction work in practice?

not sure about time to compare between sparse vector and map

The mechanics of TFIDF is unclear

COMP 4446 / 5046
Lecture 2, 2025

Core Components
The Baseline
Word Vectors
Evaluation
Workshop Preview

menti.com 7927 6661

Muddy Card

Open now, closes at 7:05pm

https://saipll.shinyapps.io/student-interface/

If you do not wish to participate in the study, use the Ed form instead

Go to Ed → Lessons → Muddy Cards Lecture 2