MANCHESTER 1824
The University of Manchester

Deep Learning Approaches for Sarcasm Detection

Student:
Miruna Pislar

Supervisor:
Mr. John McNaught

MANCHESTER 1824
The University of Manchester

Contents

- Introduction
- Background
- Sarcasm Datasets
- Evaluation metrics
- Traditional approaches: a baseline implementation
- Deep learning approaches: introduction to word2vec
- Future plans and aspirations

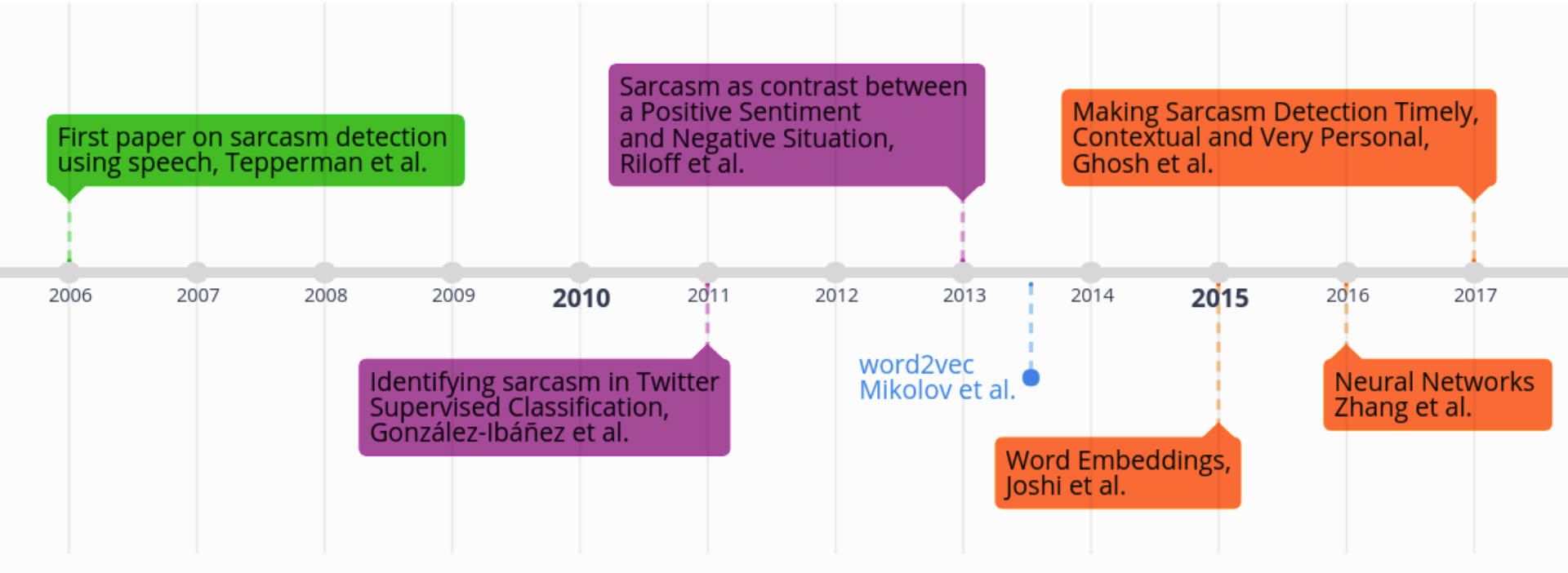
MANCHESTER 1824
The University of Manchester

Introduction

- Sarcasm is a bitter expression or remark
- Requires **wit to produce** and **wit to understand**
- Recognizing sarcastic intent in written text:
  - Interjections
  - Intensifiers
  - Punctuation
  - Capitalization
  - Hyperbole
- Accurate sarcasm detection significantly improves the performance of sentiment analysis tasks.

MANCHESTER 1824
The University of Manchester

Background



Timeline also available at https://time.graphics/line/23643
Created with TimeGraphics

MANCHESTER 1824
The University of Manchester

Sarcasm Corpus

- Unfortunately, no standard sarcastic corpus
- I use a Twitter dataset made publicly available by Ghosh *et al.*
- Retrieval cue: #sarcasm
- Training set: 39K tweets (18K sarcastic, 21K non-sarcastic)
- Test set contains 3000 tweets manually annotated by an internal team of researchers

MANCHESTER 1824
The University of Manchester

Using a Twitter Dataset

Advantages	Disadvantages
• Data has already been annotated • Previously used by other researchers - a sensible comparison can be made.	• Training set is too big to be manually checked (it will contain noise) • #sarcasm does not imply that the tweet is always sarcastic
• Corpus is big enough to perform deep learning approaches. • Data is evenly split (no skew risk).	• Deep learning models are computationally intensive. • High memory requirements.
• Twitter is actively used by people – current analysis reflects real, global usage of sarcasm. • New, sensible examples can always be added.	• 140 characters limit – conversational context is hard to gather. • Ungrammatical use of language, hash-tags and emojis are making the task even harder.

MANCHESTER 1824
The University of Manchester

Evaluation

- **Evaluation metrics:** F1-score, precision and recall.
- **Manual error analysis:** ask people to label examples misclassified by the model.
- **Human level performance:** it is not always the case that people can differentiate between sarcastic and non-sarcastic utterances (tweets are noisy, lack of context).

MANCHESTER 1824
The University of Manchester

Traditional approaches

- SVC outperforms other supervised learning algorithms tried
- Main implementation idea: sarcasm most commonly appears as a **contrast** between **positive sentiments** and **negative situations** (Riloff *et al.* 2013)

  `Off to soccer practice, could hell week be any more fun ? #sarcasm`

- But not always the case:

  `A woman needs a man like a fish needs a bicycle.`

MANCHESTER 1824
The University of Manchester

Subjectivity Lexicon

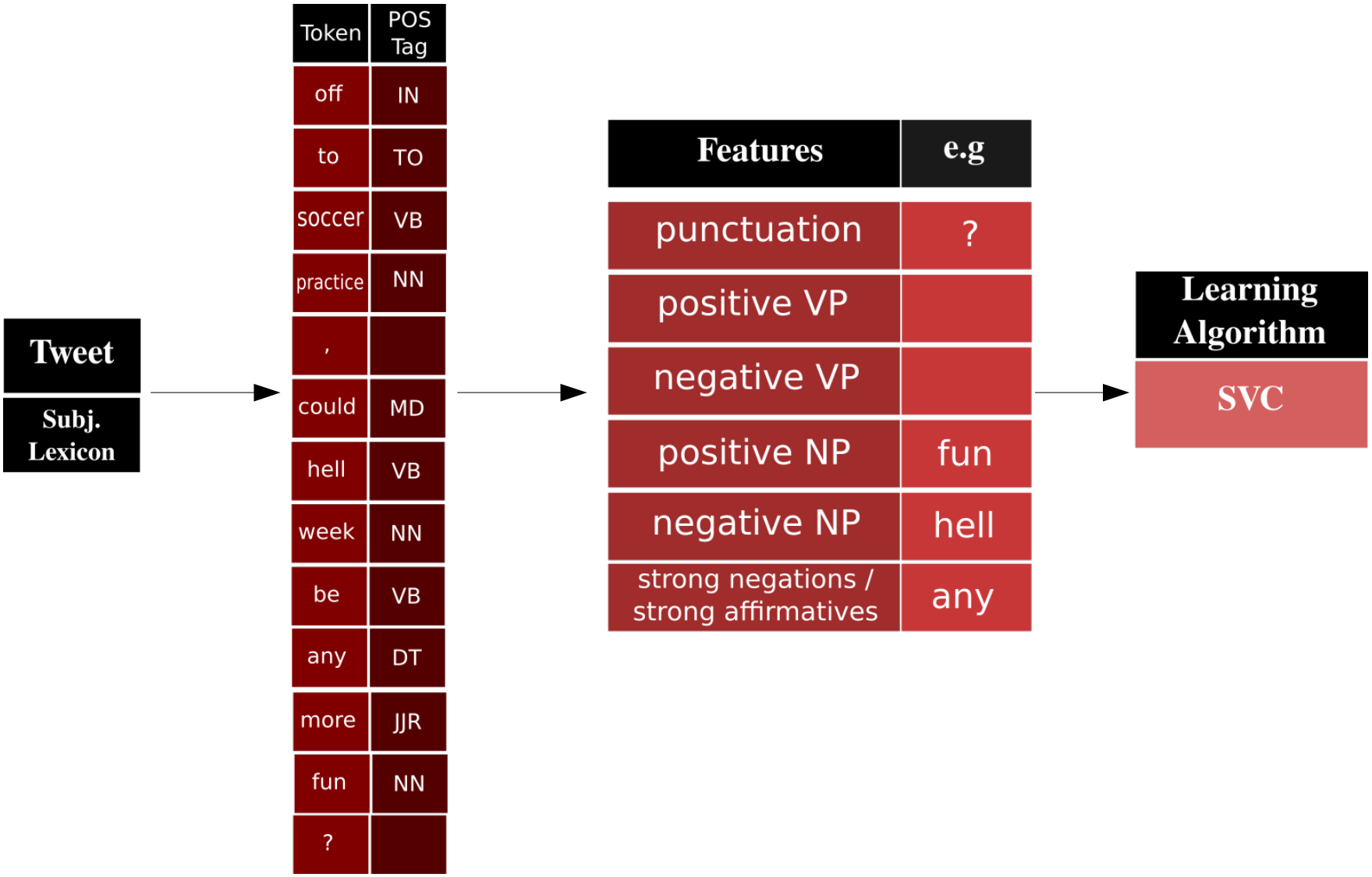
- Used to capture contrast between positive and negative phrases.

Word	Type	Part of speech	Polarity
abandoned	weak subj	adj	negative
abdicate	weak subj	verb	negative
aberration	strong subj	adj	negative
…			
zealously	strong subj	anypos	negative
zenith	strong subj	noun	positive
zest	strong subj	noun	positive

 Based on the paper: Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Theresa Wilson, Janyce Wiebe, and Paul Hoffmann 2005. *Proc. of HLT-EMNLP*-2005

MANCHESTER 1824
The University of Manchester

Feature extraction



MANCHESTER 1824
The University of Manchester

Baseline Models

Model	Features					
1	PosVP	NegVP	PosNP	NegNP	StrongN	Punct.
2	Pos1-gram	Neg1-gram	Pos/Neg	Punct.	StrongN	StrongA
3	PosNegVP	PosNegNP	Pos/Neg	Punct.	StrongN	StrongA

- **Conclusion:** finding suitable features is a tedious process.
- Human language is too complex to attempt at modeling it solely based on rules and patterns.

MANCHESTER 1824
The University of Manchester

Baseline Evaluation

	Support Vector Classifier		
	Precision	Recall	F-score
Model 1	0.59424	0.61144	0.59669
Model 2	0.53363	0.54853	0.53894
Model 3	0.60222	0.61795	**0.60450**

- **Suggestion:** use words as features (build a bag-of-words model)

MANCHESTER 1824
The University of Manchester

Bag of Words

- The quick brown fox jumps over the lazy dog
- Never jump over the lazy dog quickly

MANCHESTER 1824
The University of Manchester

Bag of Words

- The quick brown fox jumps over the lazy dog
- Never jump over the lazy dog quickly


```
{
    'brown': 0,
    'dog': 1,
    'fox': 2,
    'jump': 3,
    'jumps': 4,
    'lazy': 5,
    'never': 6,
    'over': 7,
    'quick': 8,
    'quickly': 9,
    'the': 10,
}
```

MANCHESTER 1824
The University of Manchester

Bag of Words

```
{
    'brown': 0,
    'dog': 1,
    'fox': 2,
    'jump': 3,
    'jumps': 4,
    'lazy': 5,
    'never': 6,
    'over': 7,
    'quick': 8,
    'quickly': 9,
    'the': 10
}
```

MANCHESTER 1824
The University of Manchester

Bag of Words

```
{
    'brown': 0,
    'dog': 1,
    'fox': 2,
    'jump': 3,
    'jumps': 4,
    'lazy': 5,
    'never': 6,
    'over': 7,
    'quick': 8,
    'quickly': 9,
    'the': 10
}
```

brown	dog	fox	jump	jumps	lazy	never	over	quick	quickly	the
1	1	1	0	1	1	0	1	1	0	2
0	1	0	1	0	1	1	1	0	1	1

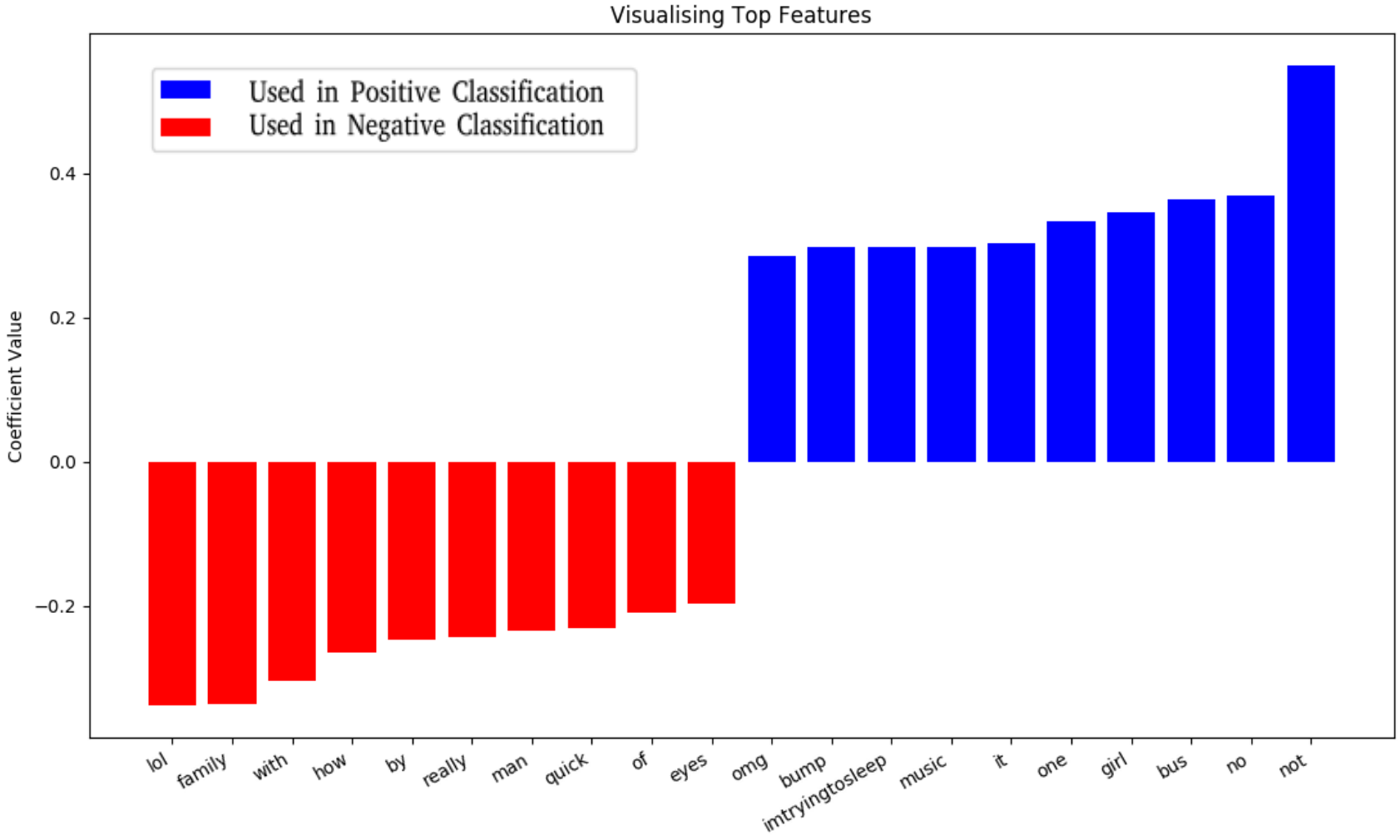
MANCHESTER 1824
The University of Manchester

Bag of Words

- Different scoring methods – count, frequency, binary, tf-idf
- Used BoW with Linear SVC
- Compared it with a Feed Forward Neural Network
- Performed feature analysis (see which are the important words)

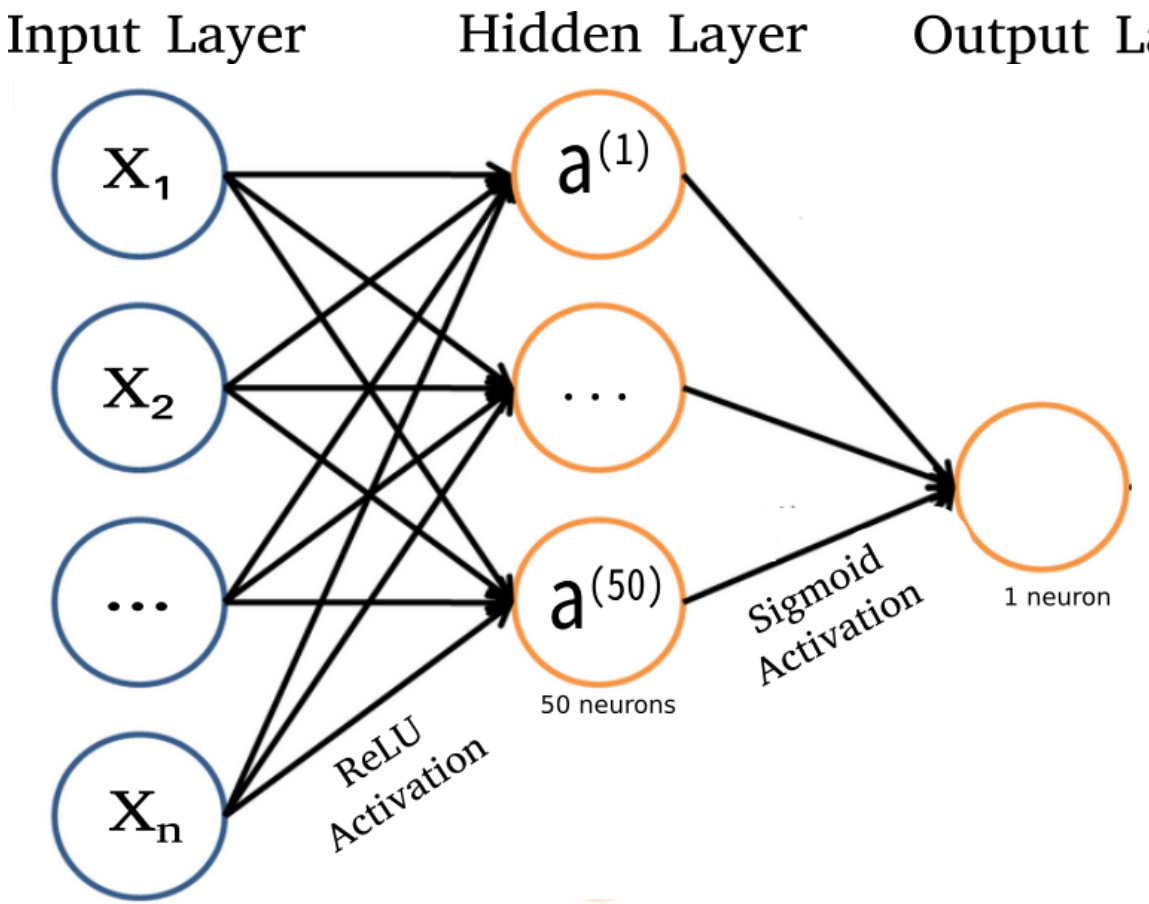
MANCHESTER 1824
The University of Manchester

Analysis of lexical features



MANCHESTER 1824
The University of Manchester

BoW with a Feed Forward Neural Network Architecture



- Mini-batch gradient descent (32 batches)
- Binary cross-entropy
- Adam Optimization
- 10 – 30 epochs
- Scoring methods: binary, count, freq, tf-idf

MANCHESTER 1824
The University of Manchester

Bag-of-Words Evaluation

	Count			Tf-idf		
	Precision	Recall	F-score	Precision	Recall	F-score
SVC Bow	0.8225	0.808	**0.8104**	0.8181	0.8066	0.8089
NN BoW	0.8450	0.8335	0.8354	0.8541	0.8462	**0.8478**

- **Problems:**

  – sparsity of the feature matrix: slow training, impractical for larger inputs

  – predictions strictly based on the lexical incongruities

- **Suggestion**: use word2vec

MANCHESTER 1824
The University of Manchester

Word2vec

- **Similar words will appear in similar contexts**
- Represent words as a distribution of probabilities:

  `vector[“computer”] = array([-0.00449447, -0.00310097,  0.02421786, …])`

- Similarity measured as cosine distance:

  `similarity(“woman”, “man”) = 0.73723527`

  `vector[“king”] - vector[“man”] + vector[“woman”] = vector[“queen”]`

Word2vec

Input:
one document

Lorem ipsum dolor sit amet, consete-tur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et

word vectors

Model:

vector space

most_similar('france'):

spain	0.678515
belgium	0.665923
netherlands	0.652428
italy	0.633130

highest cosine distance values in vector space of the nearest words

MANCHESTER 1824
The University of Manchester

What's next?

Task Name	Start	Finish	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18	W19	W20	W21	W22
Research on sarcasm and irony	18/09/17	29/09/17																						
Gain more deep learning knowledge	02/10/17	27/10/17																						
Get the dataset and process it	30/10/17	03/11/17																						
Design a learning model	06/11/17	17/11/17																						
Experiment with various features	13/11/17	01/12/17																						
Implement the detector	20/11/17	15/12/17																						
Evaluate the learning model	04/12/17	15/12/17																						
Improve the learning model based on results	29/01/18	09/03/18																						
Evaluate and compare with previous work	26/02/18	16/03/18																						
Prepare project submission	19/03/18	23/03/18																						

- Learn more Deep Learning approaches for NLP
- Implement and evaluate these approaches to accurately detect sarcasm in tweets
- Contrast and compare the models developed with current state-of-the-art techniques for sarcasm detection

MANCHESTER 1824
The University of Manchester

Conclusion

- Sarcasm is a **very subjective phenomenon -** even for human annotators it is quite hard to decide if the speaker was sarcastic or not.
- It would be interesting to contrast sarcasm with **irony** – in theory, the difference seems subtle, yet in practice people will easily distinguish between the two.
- Additionally, sarcasm is **culture-specific**. Liu *et al.* (2014) as well as Joshi *et al.* (2016) show that sarcasm depends on language traits and cultural background.
- Making a parallel between English sarcasm and the sarcasm of my native language would be an extra that I am keeping in mind in case I have enough time.

MANCHESTER 1824
The University of Manchester

References

- Joseph Tepperman, David R Traum, and Shrikanth Narayanan. 2006. ” yeah right”: sarcasm recognition for spoken dialogue systems.. In *INTERSPEECH*. Citeseer.
- Roberto González-Ibánez, Smaranda Muresan, and Nina Wacholder. 2011. Identifying sarcasm in Twitter: a closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, 581–586
- Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. Sarcasm as Contrast between a Positive Sentiment and Negative Situation. *EMNLP*. 704–714
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv:1301.3781*.
- Debanjan Ghosh, Weiwei Guo, and Smaranda Muresan. 2015a. Sarcastic or Not: Word Embeddings to Predict the Literal or Sarcastic Meaning of Words. In *EMNLP*.
- Meishan Zhang, Yue Zhang, Guohong Fu. 2016. Tweet Sarcasm Detection Using Deep Neural Network. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2449–2460

MANCHESTER 1824
The University of Manchester

References

- Armand Joulin, Edouard Grave, Piotr Bojanowski, Tomas Mikolov. 2016. Bag of Tricks for Efficient Text Classification. *arXiv:1607.01759*
- Aditya Joshi, Pushpak Bhattacharyya, Mark Carman, Jaya Saraswati, and Rajita Shukla. 2016. How Do Cultural Differences Impact the Quality of Sarcasm Annotation?: A Case Study of Indian Annotators and American Text. *LaTeCH* 2016 (2016), 95.
- Aditya Joshi, Vinita Sharma, and Pushpak Bhattacharyya. 2015. Harnessing context incongruity for sarcasm detection. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Vol. 2. 757–762.
- Aditya Joshi, Vaibhav Tripathi, Kevin Patel, Pushpak Bhattacharyya, and Mark Carman. 2016b. Are Word Embedding-based Features for Sarcasm Detection? *EMNLP 2016* (2016)
- Aniruddha Ghosh and Tony Veale. 2016. Fracking Sarcasm using Neural Network. *WASSA NAACL* 2016 (2016).
- Aniruddha Ghosh and Tony Veale. 2017. Magnets for Sarcasm: Making Sarcasm Detection Timely, Contextual and Very Personal. Conference on Empirical Methods in Natural Language Processing (*EMNLP*). 7th-11th September, 2017, Copenhagen, Denmark.

MANCHESTER 1824
The University of Manchester

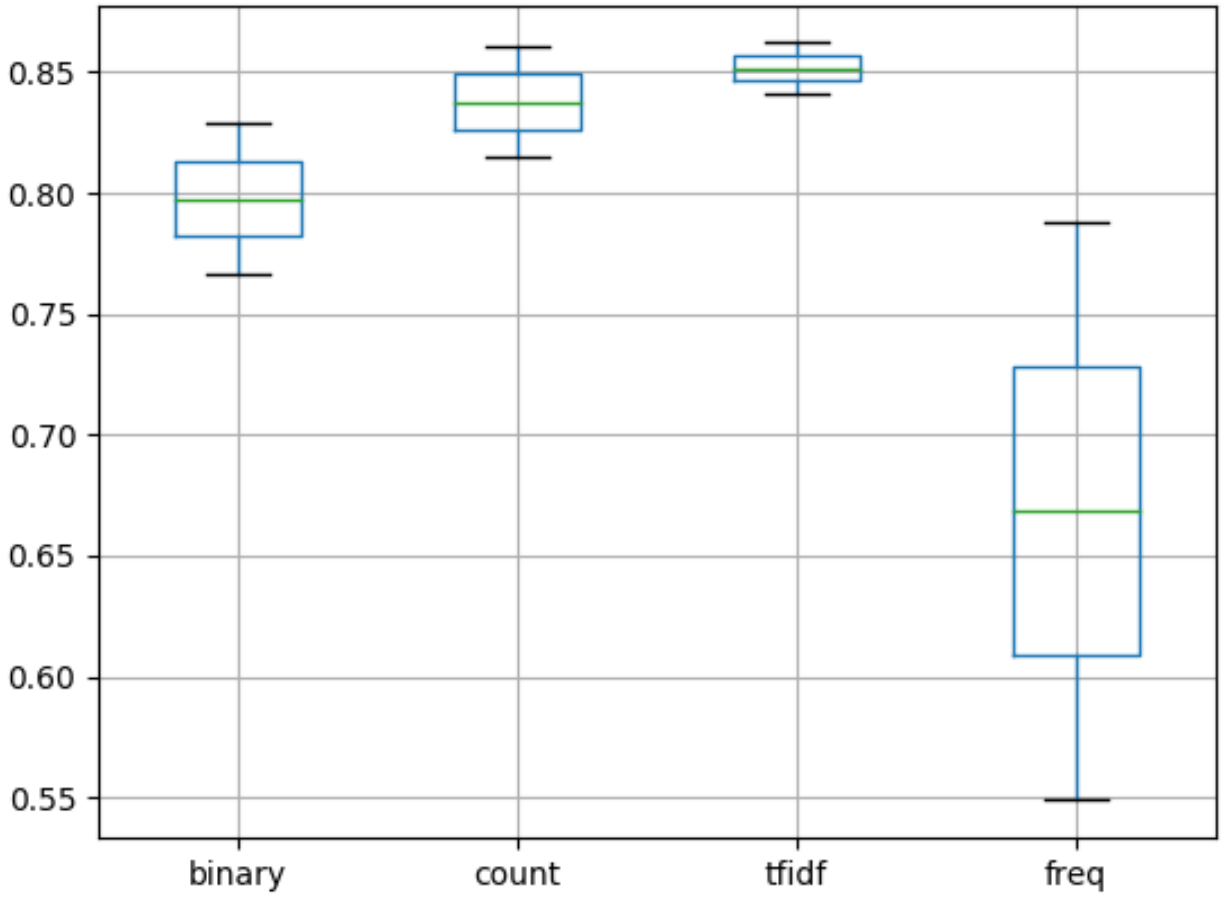
BoW Scoring Methods

- There are different methods for scoring words in a BoW model; four of tested ones are:

1. **binary** - words are marked as present (1) or absent (0).
2. **count** - the occurrence count for each word is marked as an integer.
3. **frequency** - words are scored based on their frequency of occurrence.
4. **tf-idf** - words are scored based on their frequency (but common words are penalized)

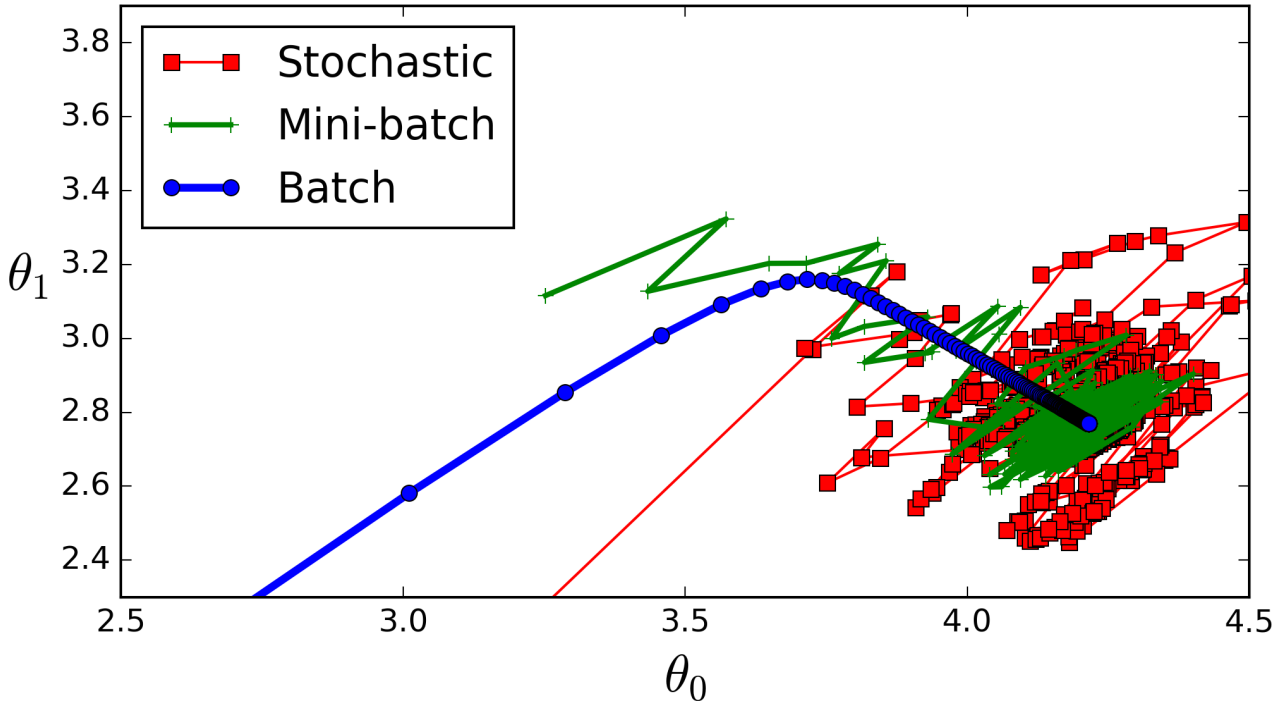
MANCHESTER 1824
The University of Manchester

BoW Scoring Methods



Choosing the batch size

- **Batch size** = number of samples that will be propagated through the network.



Advantage: it requires less memory, training is faster.

Disadvantage: the smaller the batch, the less accurate estimate of the gradient.

Source: Cross Validated - Stack Exchange.

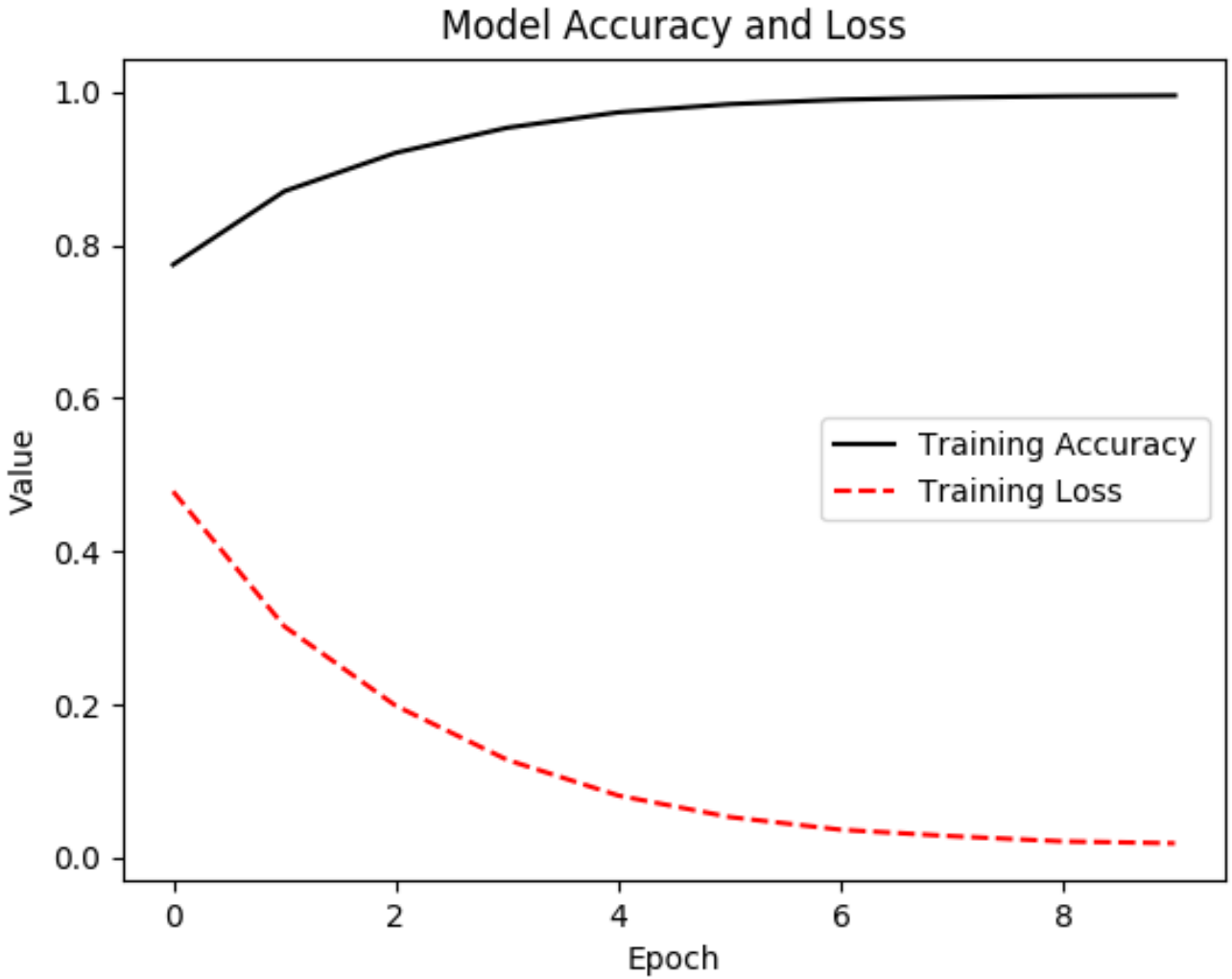
MANCHESTER 1824
The University of Manchester

Batch, epoch, iteration

- **One epoch -** one forward pass and one backward pass of **all** the training examples.
- **The batch size -** the number of training examples **in one** forward/backward **pass**. The higher the batch size, the more memory space needed.
- **Number of iterations -** number of passes, each pass using *batch size* number of examples.

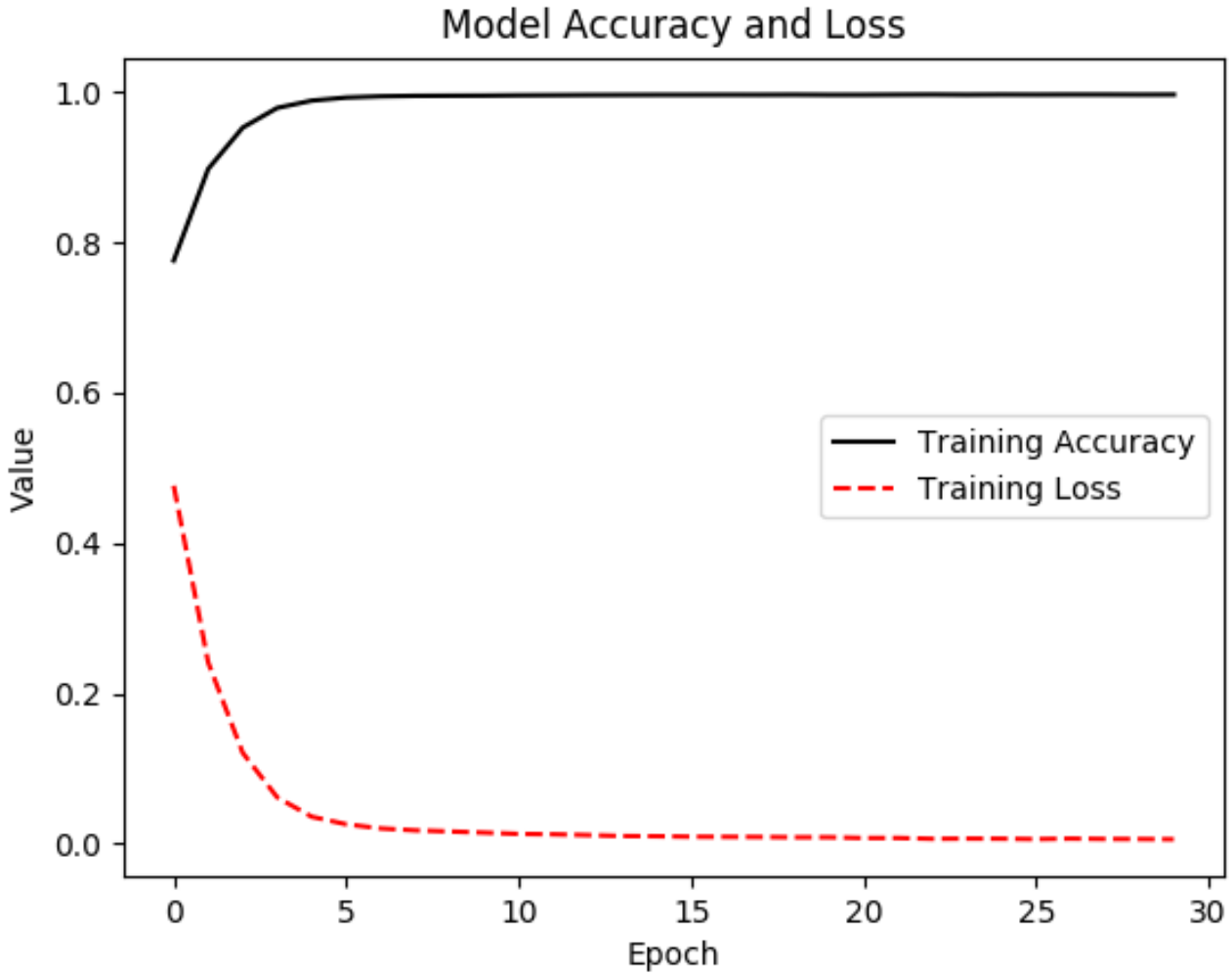
MANCHESTER 1824
The University of Manchester

Plotting the loss function



MANCHESTER 1824
The University of Manchester

Plotting the loss function



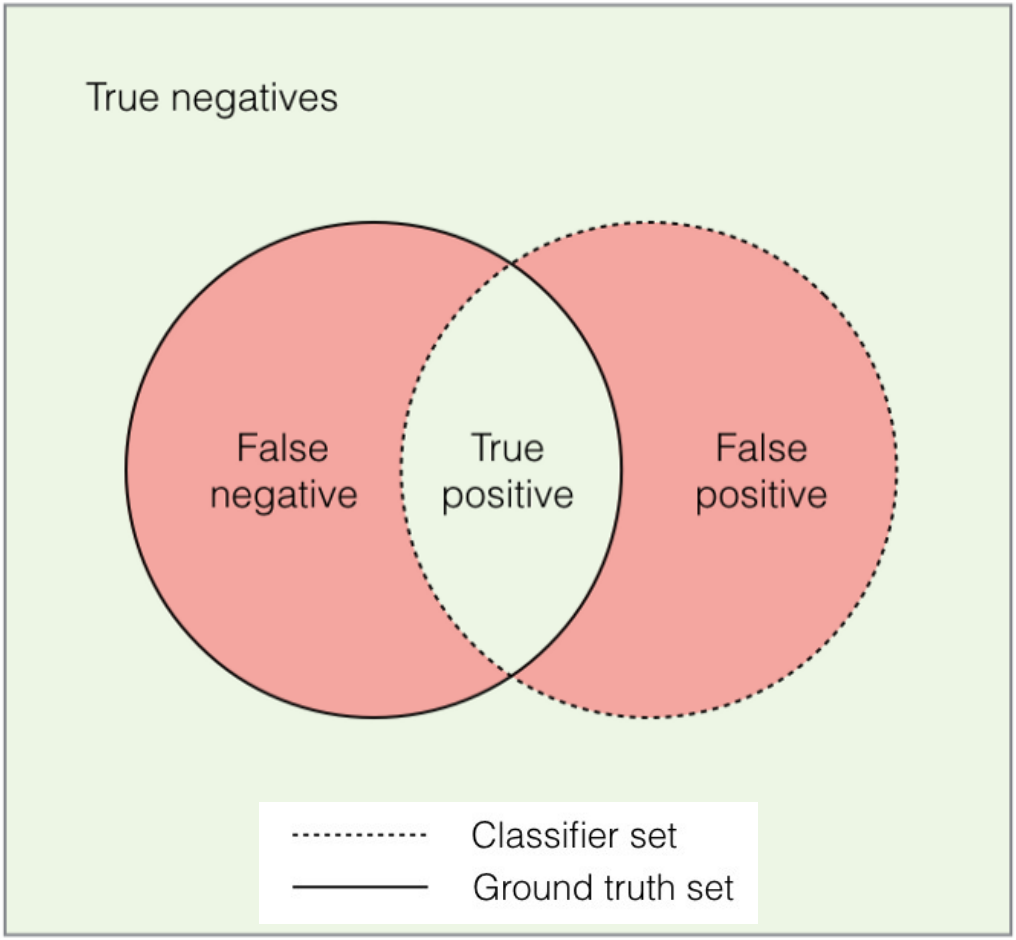
MANCHESTER 1824
The University of Manchester

Precision, Recall, F-score

- **True negatives** – samples correctly classified as not belonging to class
- **False negatives** – samples incorrectly classified as not belonging to class
- **True positives** – samples correctly classified as belonging to class
- **True positives** – samples incorrectly classified as belonging to class

- **Precision** = TP ÷ (TP + FP)
- **Recall** = TP ÷ (TP + FN)
- **F-score** = 2 * (Precision * Recall) ÷ (Precision + Recall)

Precision, Recall, F-score



Source: A Medium Corporation[US]

MANCHESTER 1824
The University of Manchester

Evaluation per class

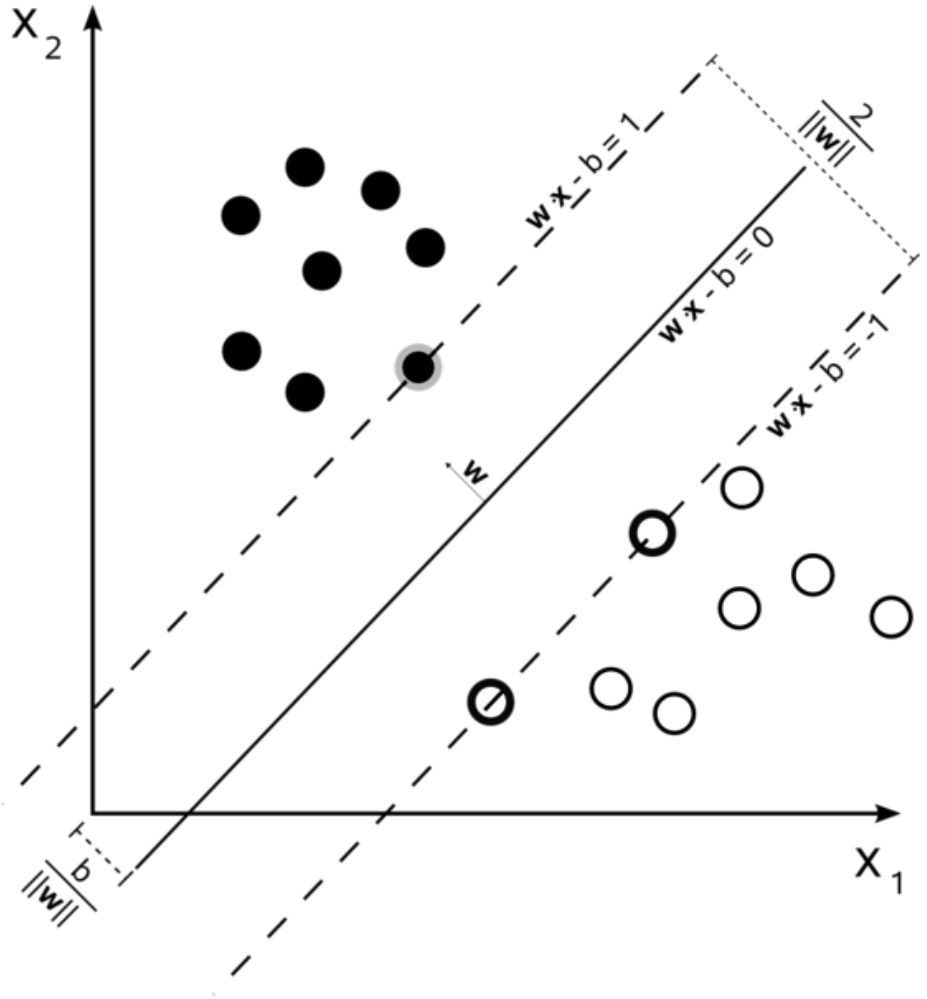
- Feed-forwards neural networks using BoW with ‘count’ scoring method, 10 epochs, 32 batch size, using ReLU for the hidden layer and sigmoid for the output layer

	Precision	Recall	F-score	Support
Non-sarcastic	0.88	0.79	0.83	2280
Sarcastic	0.71	0.83	0.77	1408
avg / total	0.82	0.81	0.81	3688

MANCHESTER 1824
The University of Manchester

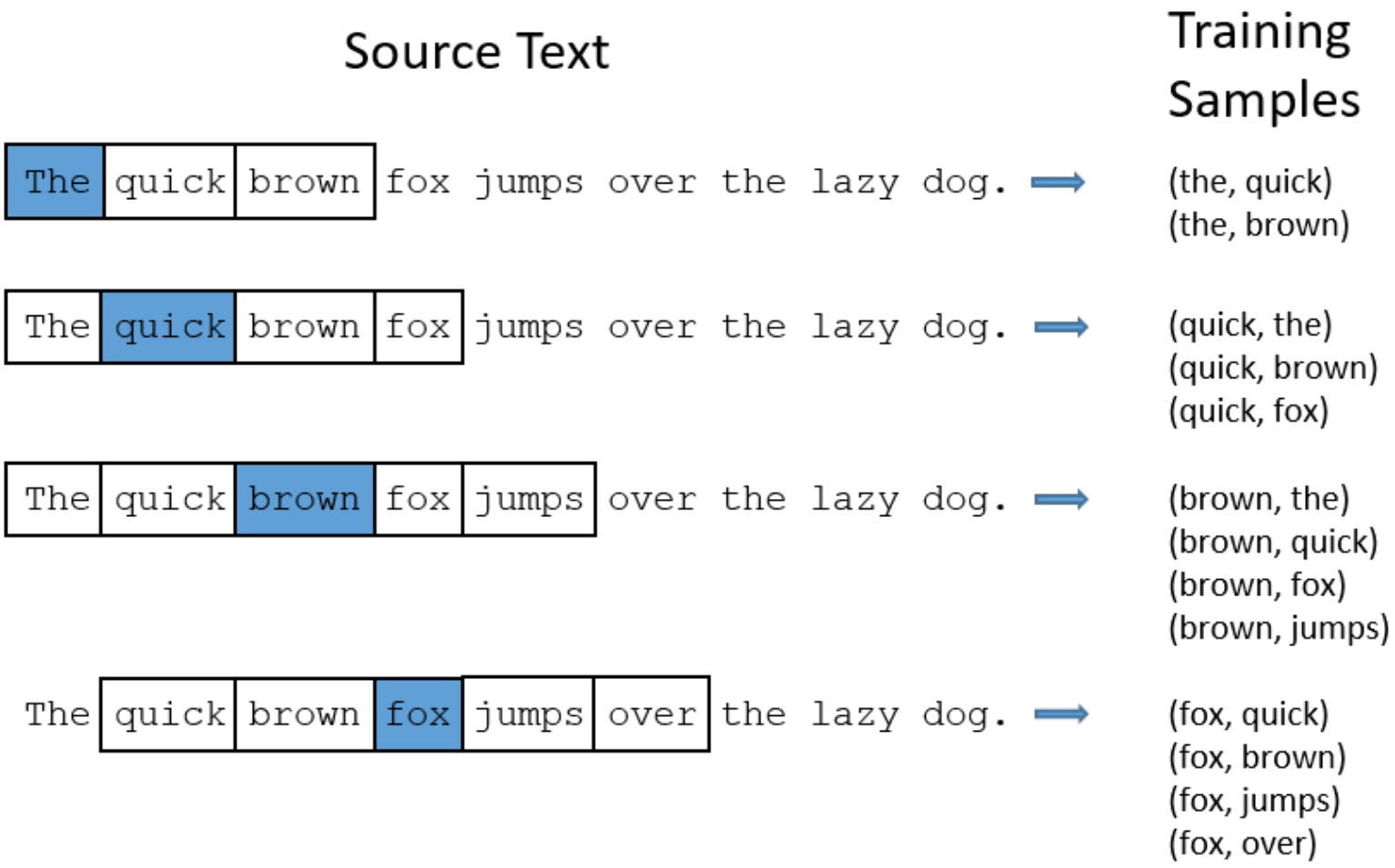
Linear Support Vector Classifier

- Has flexibility in the choice of penalties and loss functions
- Scales better to large numbers of samples
- L2 regularization penalty
- Penalty parameter C of the error term set to 1.0 (default)
- It chooses the hyperplane that represents the largest separation, or margin, between the two classes



MANCHESTER 1824
The University of Manchester

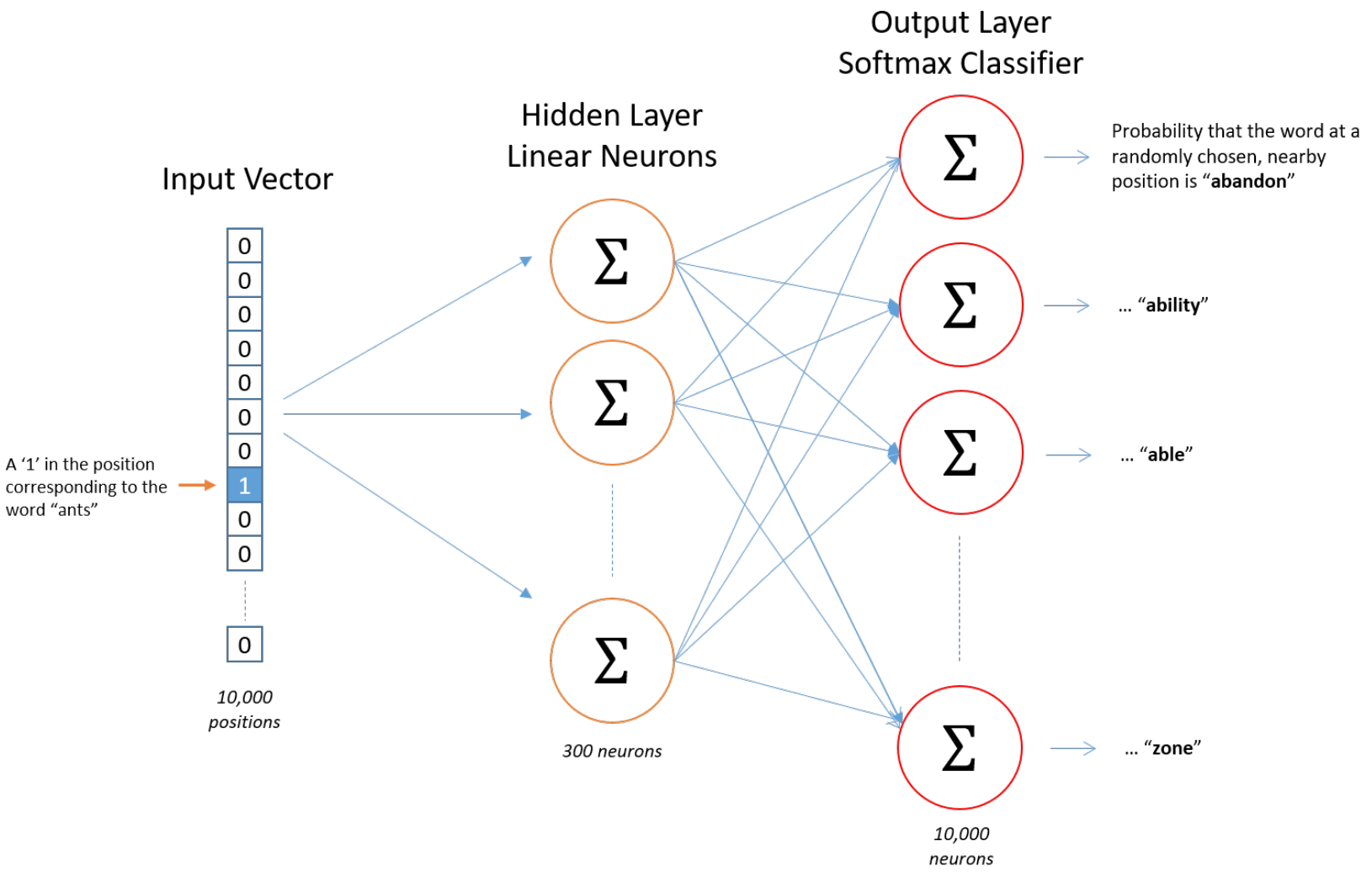
Word2vec



Source: McCormick, C. (2016, April 19). Word2Vec Tutorial - The Skip-Gram Model.

MANCHESTER 1824
The University of Manchester

Skip-gram Architecture



Source: McCormick, C. (2016, April 19). Word2Vec Tutorial - The Skip-Gram Model.