NUS National University of Singapore | Institute of Data Science

PhD-Teach-PhD Workshop

Social Media Analytics

Jun Jiang

https://www.forbes.com/sites/dangingiss/2019/04/02/8-quotes-that-define-the-state-of-social-media-today/#48feca7a6523

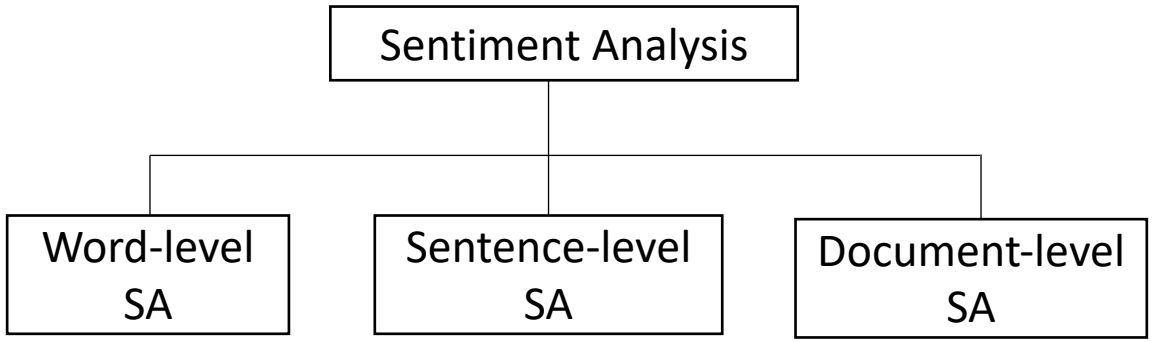
Sentiment Analysis in Social Media

Course Outline

- Introduction to SA
- Lexicon-Based Approaches
- Traditional Machine Learning Approaches
- Neural Network Approaches

What is Sentiment Analysis?

- Process of analyzing whether a text is “positive”, "negative" or "neutral".



Word-level Sentiment Analysis

What is the sentiment of ‘crude’?

- Trump makes crude remarks about immigration. (Negative)
- Crude oil is obtained by extraction from the sea beds. (Neutral)

Sentence-level Sentiment Analysis

What is the sentiment for each sentence?

- President Trump should be considered one of America's greatest presidents for his achievements so far. (Positive)
- It is difficult to imagine how he can make America great again. (Negative)

Document-level Sentiment Analysis

Opinion

Trump Is the Founders' Worst Nightmare

Once in the Oval Office, a demagogue can easily stay there.

By Bob Bauer

Mr. Bauer served as a White House counsel under President Barack Obama.

Dec. 2, 2019

2050

Donald Trump's Republican congressional allies are throwing up different defenses against impeachment and hoping that something may sell. They say that he didn't seek a corrupt political bargain with Ukraine, but that if he did, he failed, and the mere attempt is not impeachable. Or that it is not clear that he did it, because the evidence against him is unreliable "hearsay."

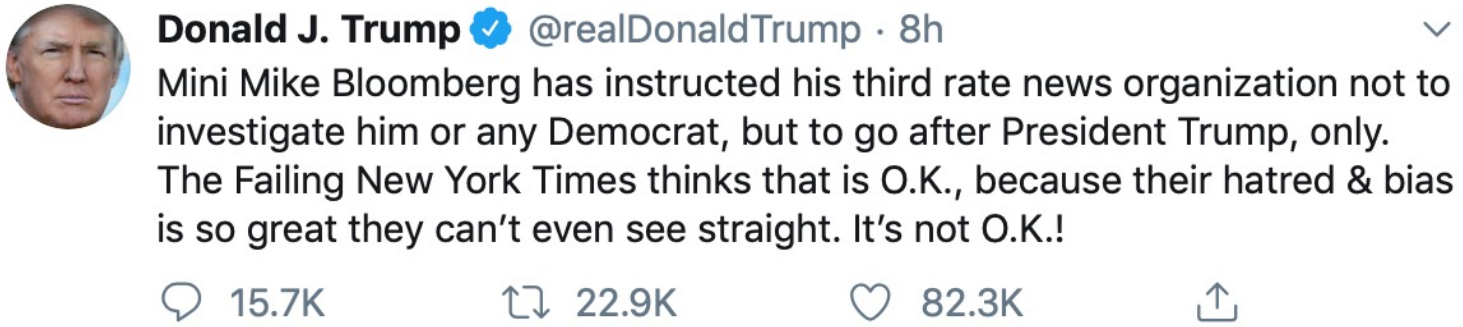
It's all been very confusing. But the larger story — the crucial *constitutional* story — is not the incoherence of the president's defense. It is more that he and his party are exposing limits of impeachment as a response to the presidency of a demagogue.

The founders feared the demagogue, who figures prominently in the Federalist Papers as the politician who, possessing "perverted ambition," pursues relentless self-aggrandizement "by the confusions of their country." The last of the papers, Federalist No. 85, linked demagogy to its threat to the constitutional order — to the "despotism" that may be expected from the "victorious demagogue." This "despotism" is achieved through systematic lying to the public, vilification of the opposition and, as James Fenimore Cooper wrote in an essay on demagogues, a claimed right to disregard "the Constitution and the laws" in pursuing what the demagogue judges to be the "interests of the people."

https://www.nytimes.com/2019/12/02/opinion/trump-impeachment.html

Sentiment Analysis in Social Media

- Posts in social media are short
- SA in Social Media is at the **sentence-level**.



Applications of Sentiment Analysis

- Allows company to gain insights to newly-launched products or promotions, etc.

Applications of Sentiment Analysis

- Brand analysis
- Reputation management

What do people think about Huawei?

Applications of Sentiment Analysis

- Understand public opinion

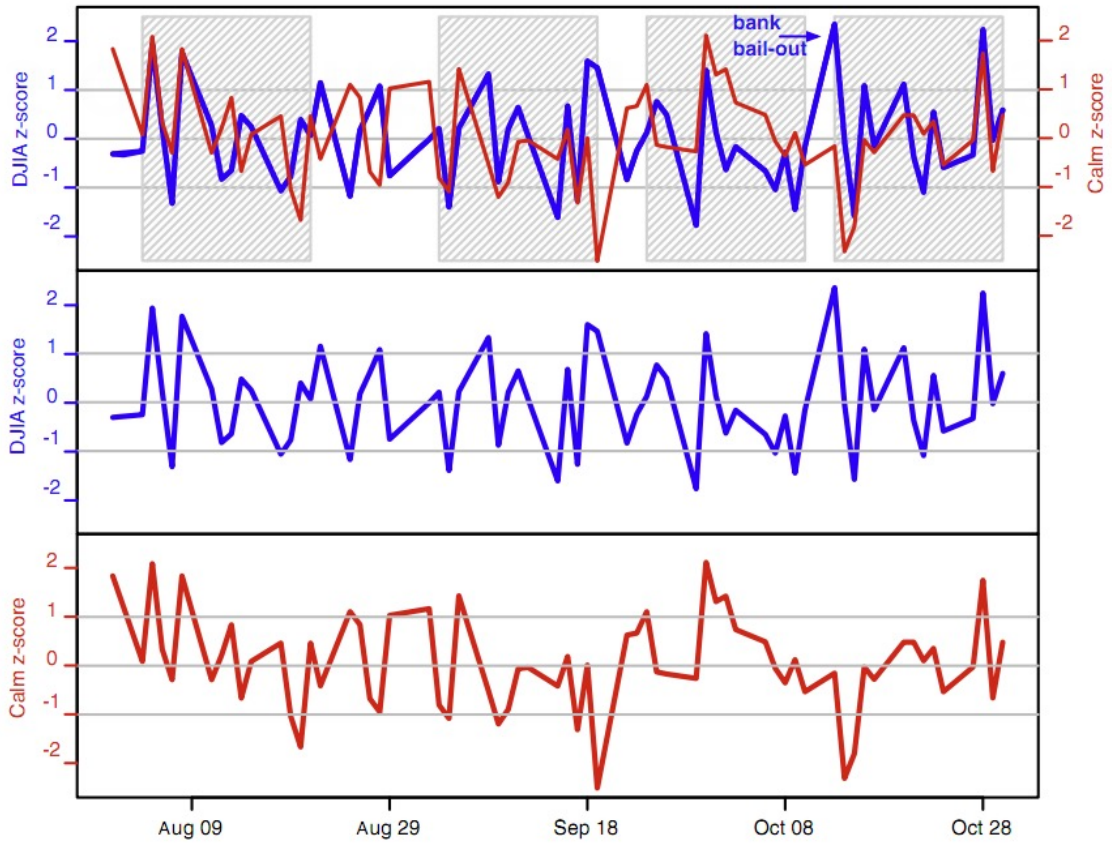
How do people view the ongoing HK protests?

Applications of Sentiment Analysis

- Financial prediction

Sentiment analysis has been

used to predict Dow Jones Index



Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market." *Journal of computational science* 2.1 (2011)

Sentiment Analysis is difficult

- **Domain dependence**

  The same sentence or phrase can have different meanings in different domains.

- **Example:**

  Which movie has the most *unpredictable* ending? (Positive)

  The vehicle's response to steering input is *unpredictable*. (Negative)

Vishal A. Kharde, S.S. Sonawane, Sentiment Analysis of Twitter Data: A Survey of Techniques , International Journal of Computer Applications, 2016

Sentiment Analysis is difficult

- **Negation**

  Sentiment can be negated in many ways other than using simple “no”, “not”, “never”, etc.

- **Example**

  “It ***avoids*** all suspense and predictability found in Hollywood movies.”

Vishal A. Kharde, S.S. Sonawane, Sentiment Analysis of Twitter Data: A Survey of Techniques , International Journal of Computer Applications, 2016

Sentiment Analysis is difficult

- **Sarcasm**

  Sarcasm typically expresses negative opinions using positive words.

- **Example**

  “Trump got a good review from his handler at the Kremlin.”

Vishal A. Kharde, S.S. Sonawane, Sentiment Analysis of Twitter Data: A Survey of Techniques , International Journal of Computer Applications, 2016

Sentiment Analysis is difficult

- **Entity dependence**

  Need to analyse the sentence structure to understand the correct sentiment for an entity

- **Example**

  A is better than B

  B is better than A.

Vishal A. Kharde, S.S. Sonawane, Sentiment Analysis of Twitter Data: A Survey of Techniques , International Journal of Computer Applications, 2016

SA in Social Media

- **Misspelled words**
  - thx → thanks
  - gr8 → great
  - l8 → late
  - Coooooool → cool
- **Acronyms**
  - GTG → good to go
  - HBD → happy birthday
  - HTH → happy to help
- **Slangs**
  - Salty: Angry or bitter about something.
  - On point: Outstanding, perfectly executed.
  - Basic: describing something that's very common, slightly negative

SA in Social Media

- **Emoticons & Emoji**

smile_positive	smile_negative
0:-)	>:(
:)	;(
:D	>:)
:*	D:<
:o	:(
:P	:|
;)	>:/

Angiani, Giulio, et al. "A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter." KDWeb. 2016.

Before introduction to methods

- A lot of challenges in SA remain to be solved
- Sentiment Analysis techniques is still developing
- There is not standard method, most of the time you need to design your model for your own problem

Lexicon-Based Approaches

Lexicon-Based Approaches

- Assume words have a prior polarity
- Use a sentiment dictionary to determine the polarity of the words in a text.
- Adjectives is a primary source of sentiment.

Taboada, Maite, et al. "Lexicon-based methods for sentiment analysis." Computational linguistics 37.2 (2011): 267-307.

Nouns, verbs & adverbs

- **Nouns, verbs and adverbs can also convey sentiment polarity.**

Examples of words in the noun and verb dictionaries.

Word	SO Value
monstrosity	−5
hate (noun and verb)	−4
disgust	−3
sham	−3
fabricate	−2
delay (noun and verb)	−1
determination	1
inspire	2
inspiration	2
endear	3
relish (verb)	4
masterpiece	5

Taboada, Maite, et al. "Lexicon-based methods for sentiment analysis." Computational linguistics 37.2 (2011): 267-307.

Intensification

- **Intensification of the polarity of words.**

e.g. slightly good. (1−50%)∗3=1.5

Percentages for some intensifiers.

Intensifier	Modifier (%)
slightly	−50
somewhat	−30
pretty	−10
really	+15
very	+25
extraordinarily	+50
(the) most	+100

Taboada, Maite, et al. "Lexicon-based methods for sentiment analysis." Computational linguistics 37.2 (2011): 267-307.

Negation

- **Negation**

1. Reverse the polarity of the word next to a negator, e.g.

good (+3) → not good (−3).

excellent (+5) → not excellent (−5) ?

2. Shift the sentiment value toward the opposite polarity by a fixed amount, e.g.

"She's not terrific (5 − 4 = 1) but not terrible (−5 + 4 = −1) either."

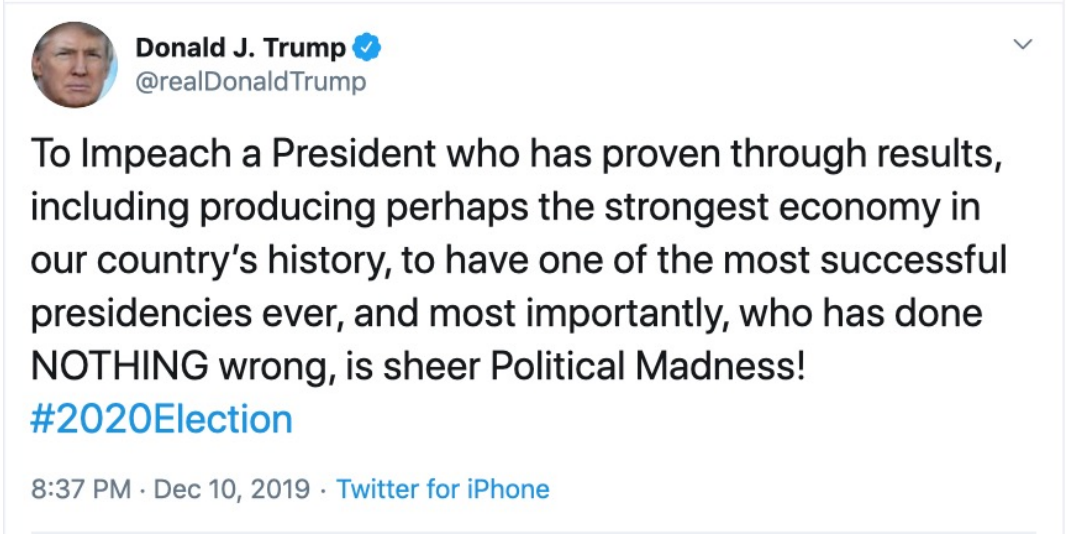
Taboada, Maite, et al. "Lexicon-based methods for sentiment analysis." Computational linguistics 37.2 (2011): 267-307.

Example

To **Impeach(-2)** a President who has proven through results, including producing perhaps the **strongest(+3)** economy in our country's history, to have one of the **most(+100%) successful(+3)** presidencies ever, and most importantly, who has done **NOTHING(negation) wrong(-3),** is sheer Political **Madness(-4)**!

$$Overall\ Sentiment$$

$$= \frac{[-2 + 3 + (1 + 100\%) * 3 \ + (-3 + 4) + (-4)]}{5} = 0.8$$



Taboada, Maite, et al. "Lexicon-based methods for sentiment analysis." Computational linguistics 37.2 (2011): 267-307.

Disadvantages

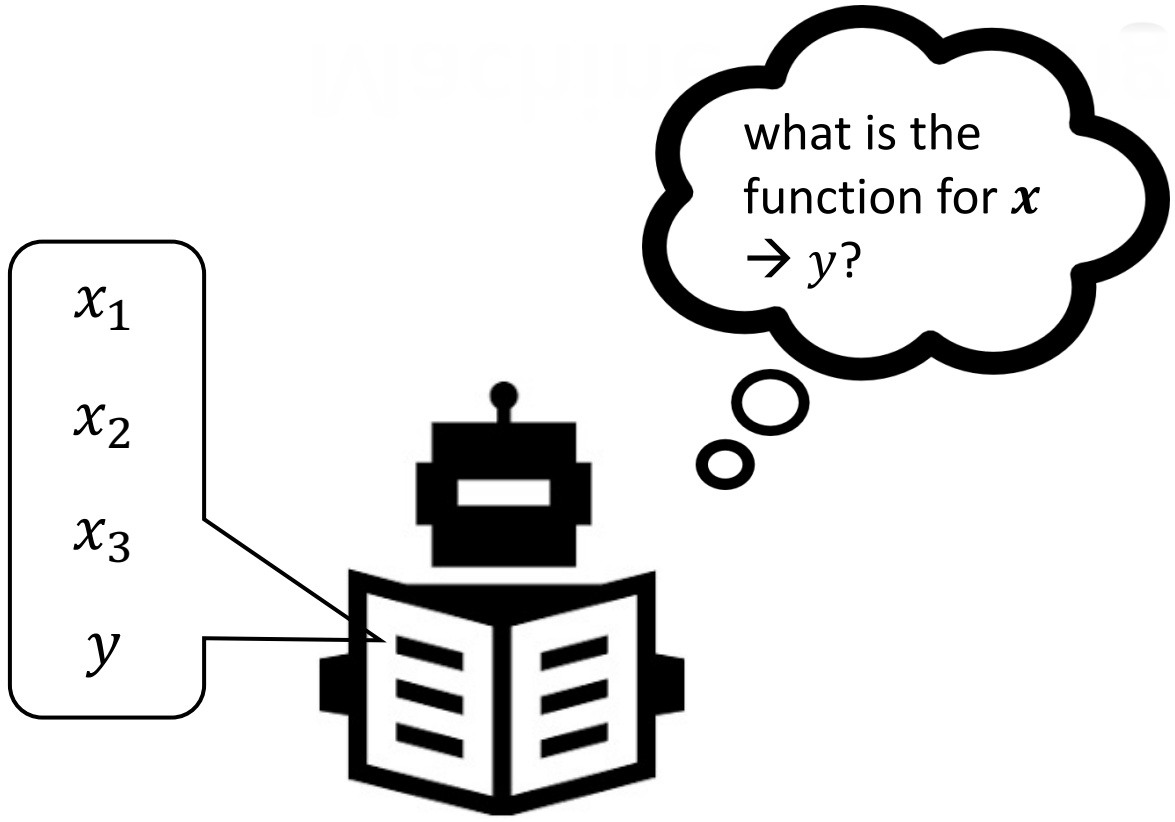
1. Manual rules are often less flexible.
2. Rules are static, but language is always evolving.
3. Relatively low accuracy.

Traditional Machine Learning Approaches

Machine Learning

Training set

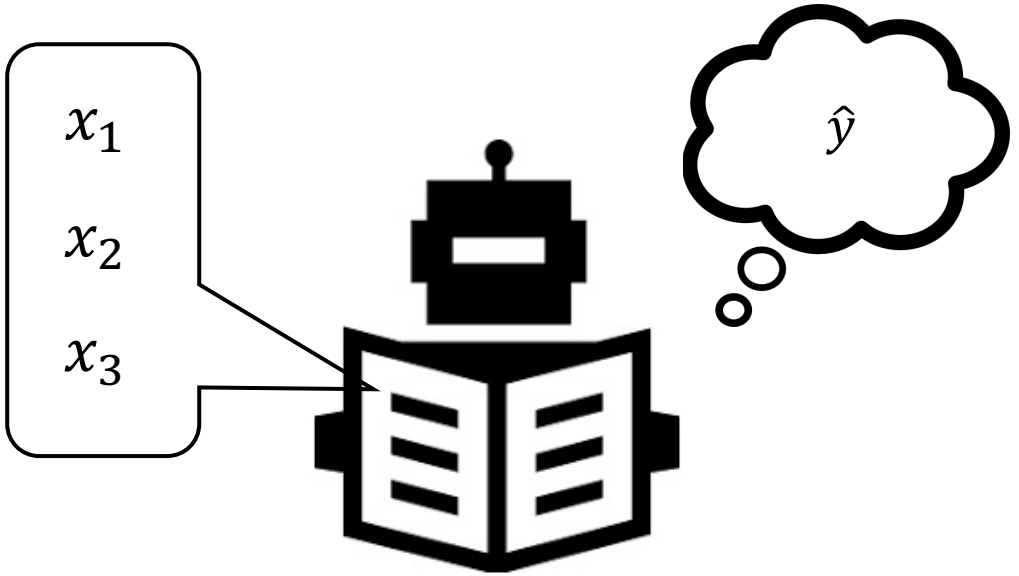
	x_1	x_2	x_3	y
Data Point 1	3	20	9	1
Data Point 2	9	18	7	1
Data Point 3	20	9	10	-1
…	…	…	…	…



Machine Learning

Test set

	x_1	x_2	x_3	y	$\hat{y}$
Data Point 1	7	10	19	1	-1
Data Point 2	4	16	6	-1	-1
Data Point 3	9	8	10	1	1
…	…	…	…	…	…



In training set & test set, we have both X & y

Machine Learning

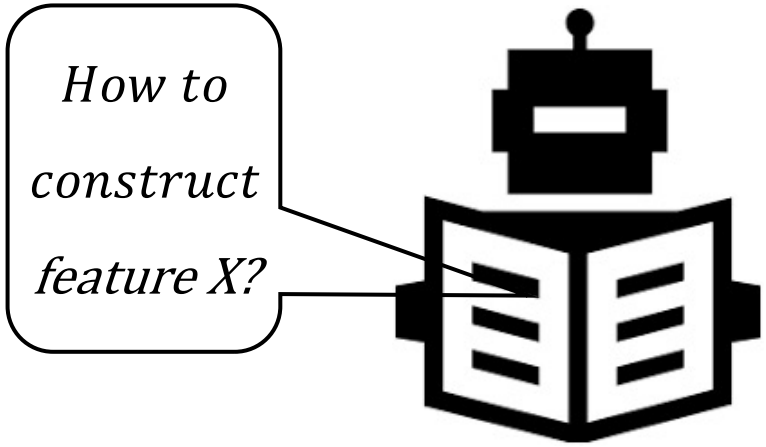
Sentence 1: Trump and Putin are very very very good friends.

Sentence 2: Good work Donald Trump.

Sentence 3: Trump is a racist.

Corpus

	x_1	x_2	…	y
Sentence 1				-1
Sentence 2				1
Sentence 3				-1



$y = 1$, the sentiment is positive

$y = -1$, the sentiment is negative

Feature Extraction

Sentence 1: Trump and Putin are very very very good friends.

Sentence 2: Good work Donald Trump.

Sentence 3: Trump is a racist.

Unigram

Trump | and | Putin | are | very | very | very | good | friends

Good | work | Donald | Trump

Trump | is | a | racists

Feature Extraction

Sentence 1: Trump and Putin are very very very good friends.

Sentence 2: Good work Donald Trump.

Sentence 3: Trump is a racist.

Unigram raw count

	Donald	Trump	and	Putin	are	very	good	friend	work	is	a	racist
Sentence 1	0	1	1	1	1	3	1	1	0	0	0	0
Sentence 2	1	1	0	0	0	0	1	0	1	0	0	0
Sentence 3	0	1	0	0	0	0	0	0	0	1	1	1

Feature Extraction

Unigram raw count

	Donald	Trump	and	Putin	are	very	good	friend	work	is	a	racist	sentiment
Sentence 1	0	1	1	1	1	3	1	1	0	0	0	0	-1
Sentence 2	1	1	0	0	0	0	1	0	1	0	0	0	1
Sentence 3	0	1	0	0	0	0	0	0	0	1	1	1	-1

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	y
Data point 1	0	1	1	1	1	3	1	1	0	0	0	0	-1
Data point 2	1	1	0	0	0	0	1	0	1	0	0	0	1
Data point 3	0	1	0	0	0	0	0	0	0	1	1	1	-1

Feature Extraction

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	y
Data point 1													
Data point 2													
Data point 3													

Training Set

Data point 1 Data point 2

……

Test Set

Data point 3

……

Feature Extraction

How to improve?

We can augment features with Bigram, Trigram, …, N-gram, to capture some space information

Sentence 1: Trump and Putin are very very very good friends.

Sentence 2: Good work Donald Trump.

Sentence 3: Trump is a racist.

Bigram

Trump and | and Putin | Putin are | are very | very very | very very | very good | good friends

Good work | work Donald | Donald Trump

Trump is | is a | a racists

Trigram, 4-gram, …, N-gram, …

Feature Extraction

Is it always better include more N-grams?

Can lead to over-fitting, i.e. you can always increase the performance in training set but may decrease performance in the test set.

How large N should be?

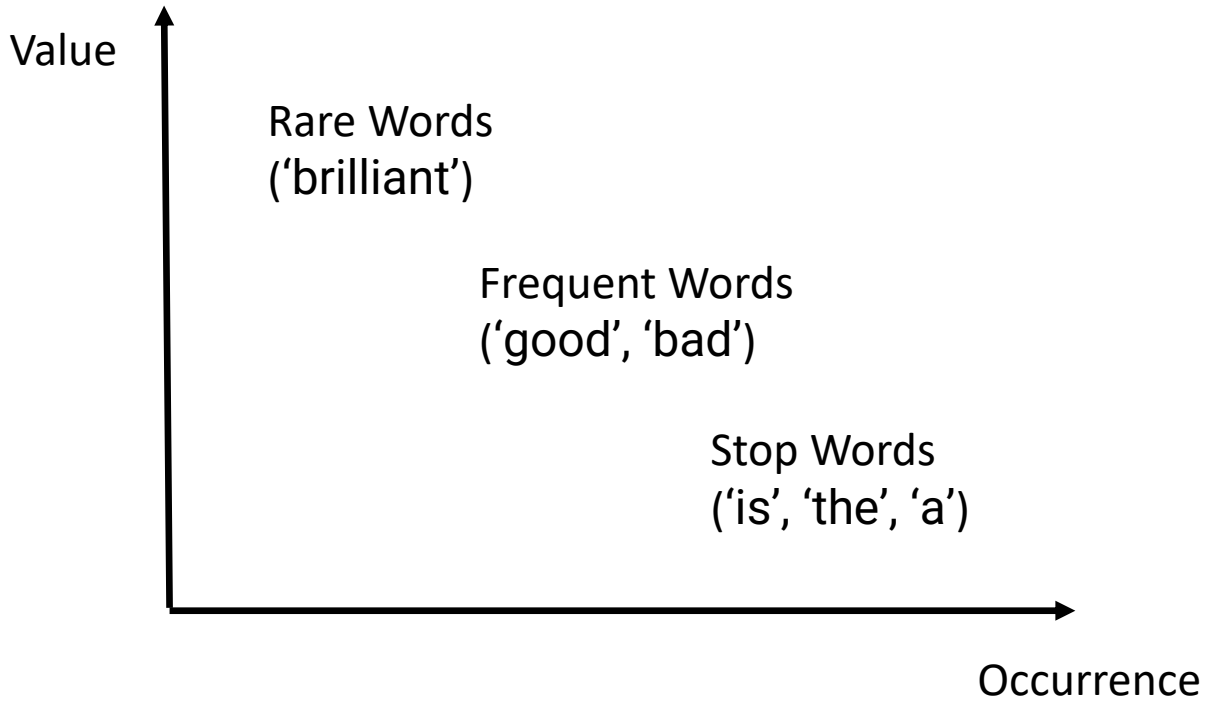
- Based on your understanding about the problem
- By trial and error.

Feature Extraction

Other than raw count, is there any other metric?

Which words are more informative?

Very common words	‘is’, ‘the’, ‘a’
Less common words	‘good’, ‘bad’
Rare words	‘brilliant’



Feature Extraction

TF-IDF:

$$\text{tfidf}(\text{t}, \text{d}, \text{D}) = tf(t, d) * idf(t, D)$$

TF:
Term Frequency

$$\text{tf}(\text{t}, \text{d}) = \frac{n_{t,d}}{n_d}$$

$n_{t,d}$: raw count of word t in sentence d

n_d : total number of words in sentence d

IDF:
Inverse Document Frequency

$$\text{idf}(t, D) = log \frac{N}{|\{d \in D : t \in d\}|}$$

N: total number of sentences in the corpus D

$|\{d \in D : t \in d\}|$: number of sentences where the word t appears

Penalising common words by assigning them lower weights

Feature Extraction

Sentence 1: Trump and Putin are very very very good friends.

Sentence 2: Good work Donald Trump.

Sentence 3: Trump is a racist.

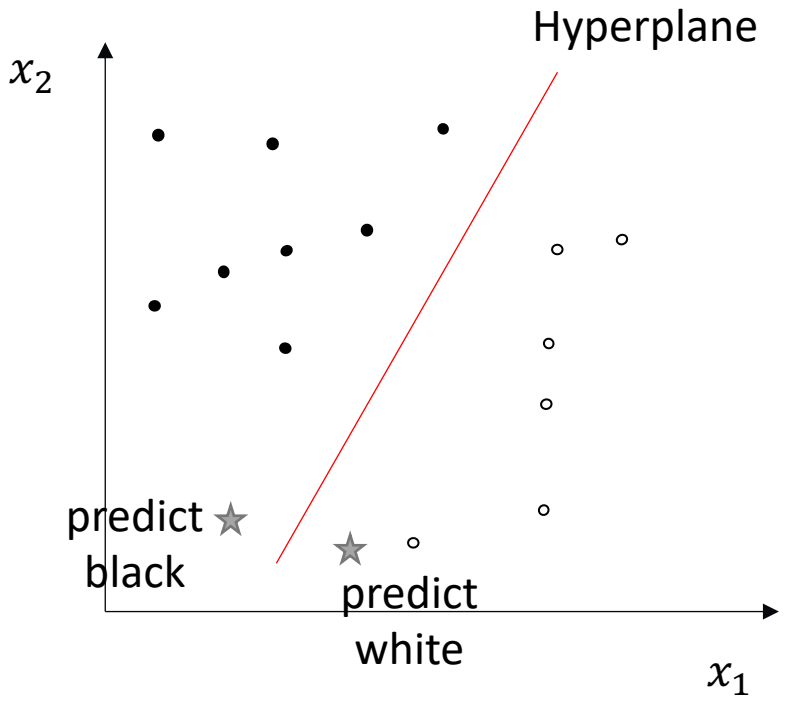
Unigram TFIDF

	Donald	Trump	and	Putin	are	very	good	friend	work	is	a	racist
Sentence 1						f_1						
Sentence 2		b_1										
Sentence 3												

$$b_1 = TF(Trump, Sentence\ 1) * IDF(Trump, Corpus) = \frac{1}{4} * \log\frac{3}{3} = 0$$

$$f_1 = TF(Very, Sentence\ 1) * IDF(Very, Corpus) = \frac{3}{9} * \log\frac{3}{1} = 0.36$$

Classification

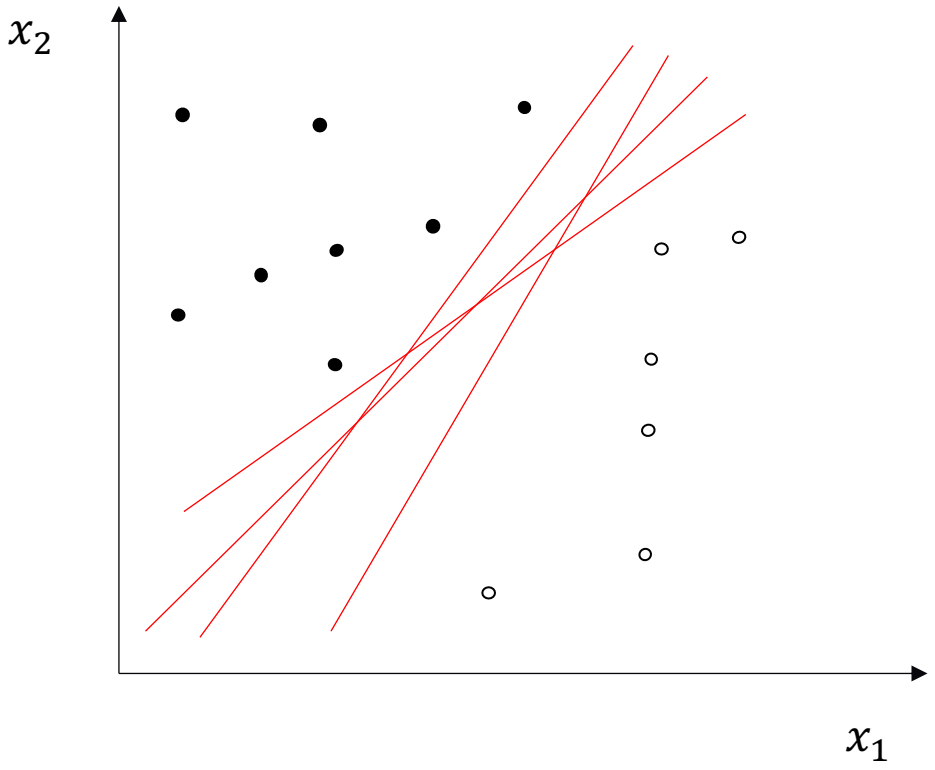


- Each point is a sentence
- Color (black/white) stands for pre-labeled sentiment value y (-1/1), i.e. negative/positive
- Use 2 features for the simplicity of illustration, e.g.

  x_1 may stand for 'raw count of *Trump*'

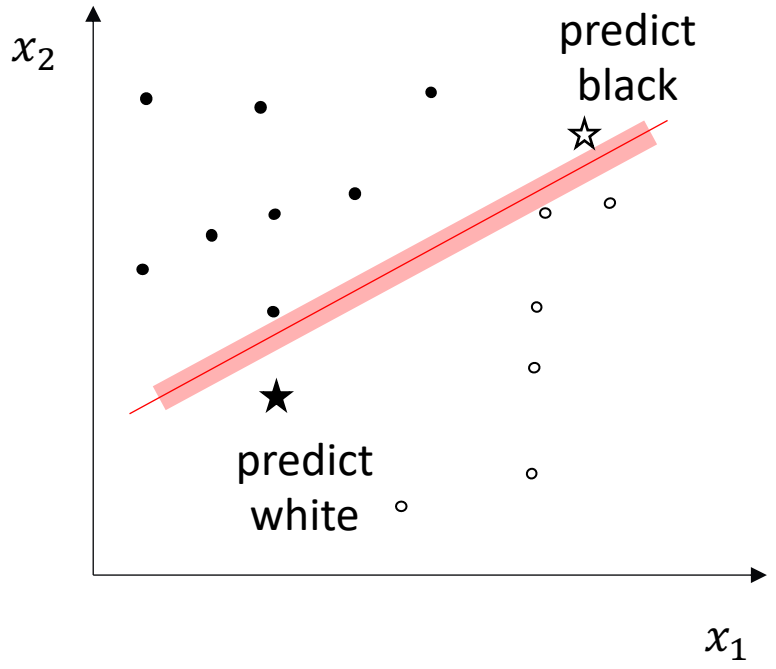
  x_2 may stand for 'raw count of *Putin*'

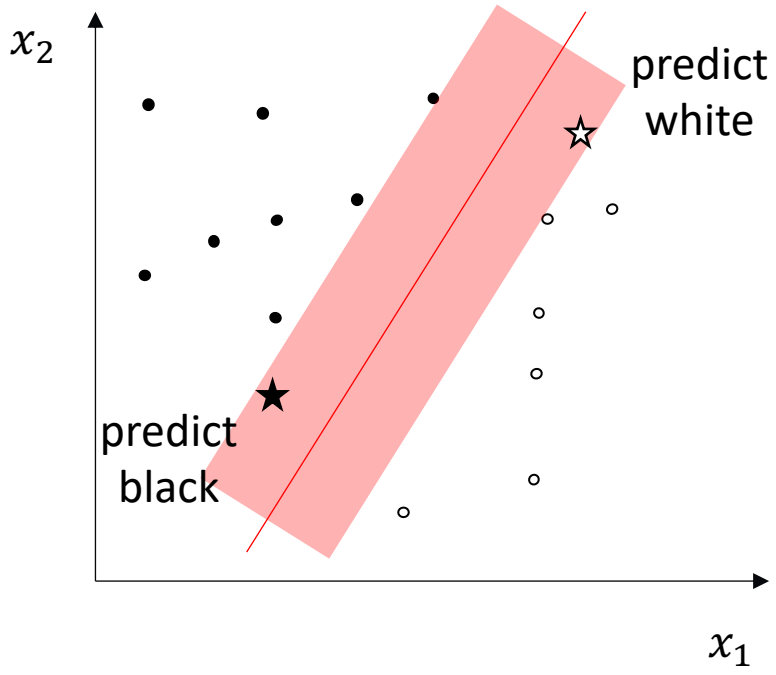
Support Vector Machine



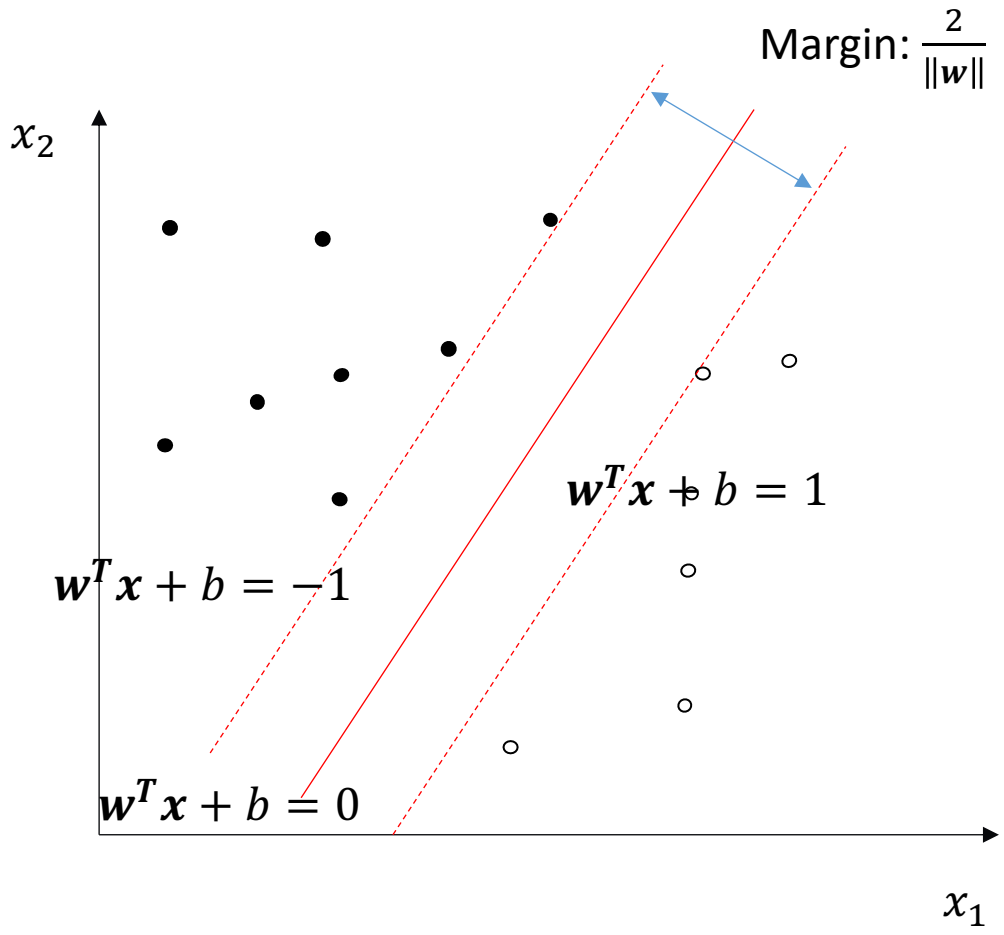
All these lines can work, but which one is the best?

Support Vector Machine





Support Vector Machine

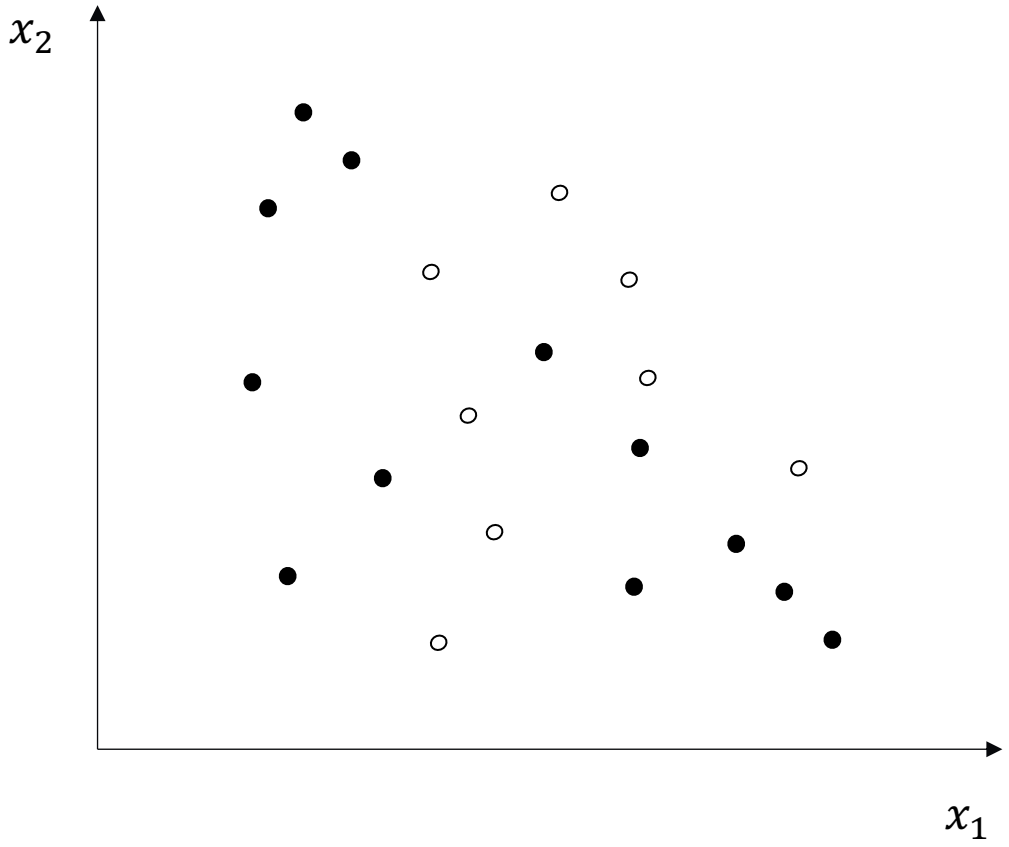


$$\max_{\mathbf{w},b} \quad \frac{2}{||\mathbf{w}||}$$
$$\text{s.t.} \quad y_i \left(\mathbf{w}^T \mathbf{x}_i + b\right) \geq 1, i = 1, 2, .., n$$

$$\Updownarrow$$

$$\min_{\mathbf{w},b} \quad \frac{||\mathbf{w}||}{2}$$
$$\text{s.t.} \quad y_i \left(\mathbf{w}^T \mathbf{x}_i + b\right) \geq 1, i = 1, 2, .., n$$

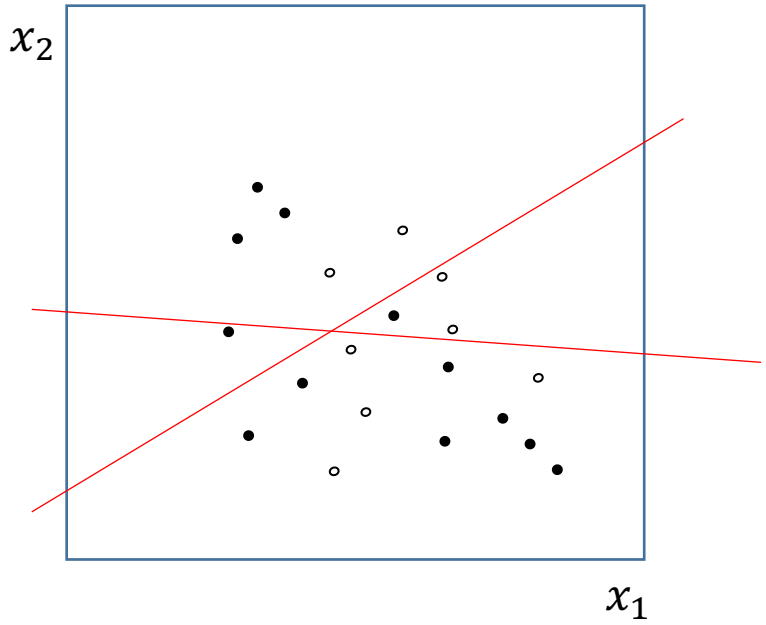
Support Vector Machine

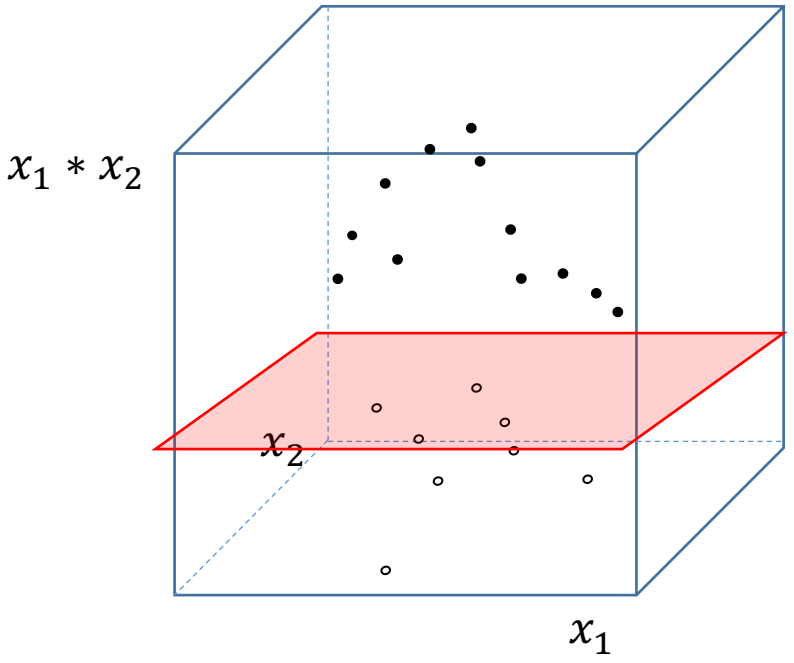


What if the data points are not separable?

Support Vector Machine

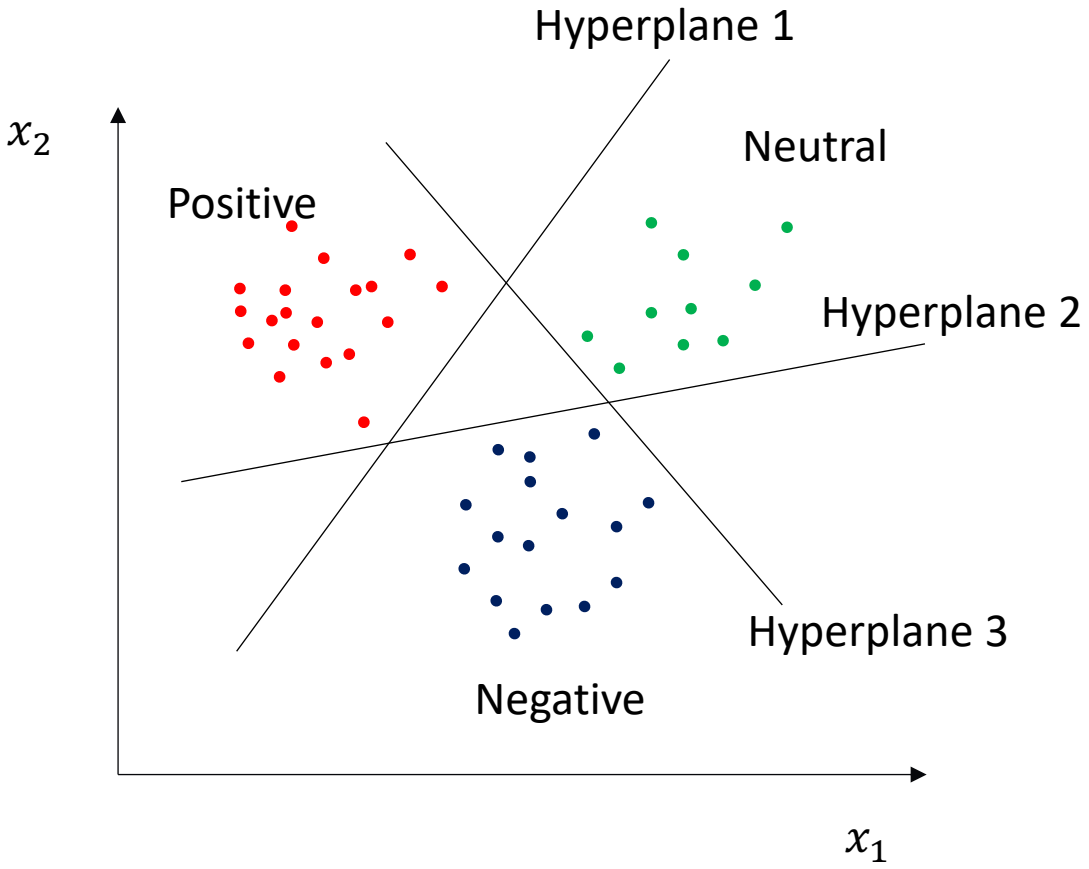
- Map data to higher dimensional space, even infinite dimensional space
- Kernel methods are used to accelerate the computation





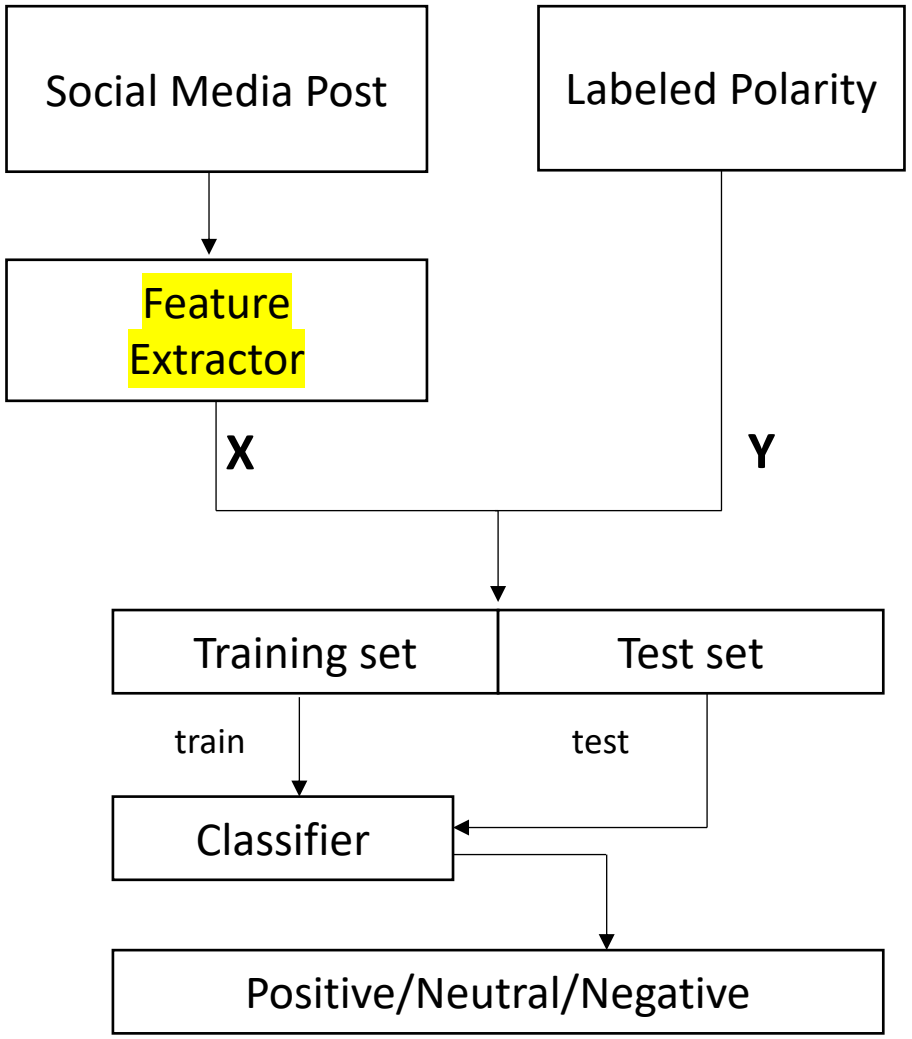
Support Vector Machine

Our sentiment label is ‘positive’/‘negative’/‘neutral’. How to perform the 3-class classification?



One-over-rest Method

Limitations



What is the limitations?

Limited feature extraction methods, cannot utilize sentence information to the most.

Neural Network Approaches

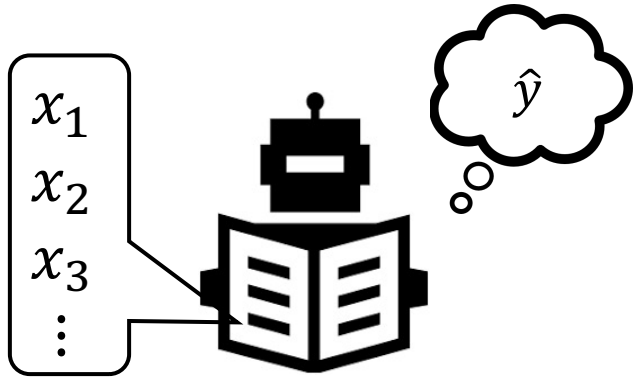
Neural Network Approaches

Sentence 1: Trump and Putin are very very very good friends.

Sentence 2: Good work Donald Trump.

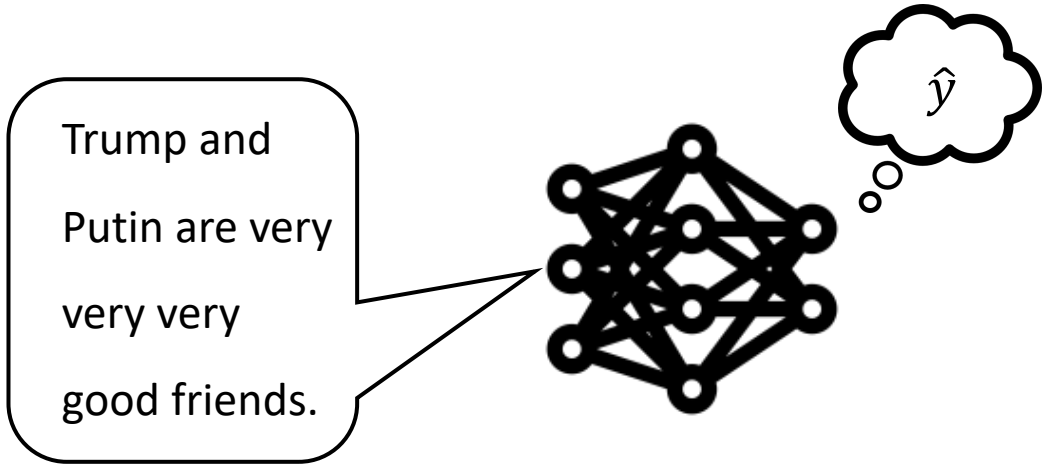
Sentence 3: Trump is a racist.

Traditional Machine Learning Approach



x_1: raw count of 'Trump' in a sentence
x_2: raw count of 'and' in a sentence
x_3: raw count of 'Putin' in a sentence
...

Neural Network Approach



Neural Network Approaches

Computer understands only numbers and do not understand words.

How to represent word with numbers?

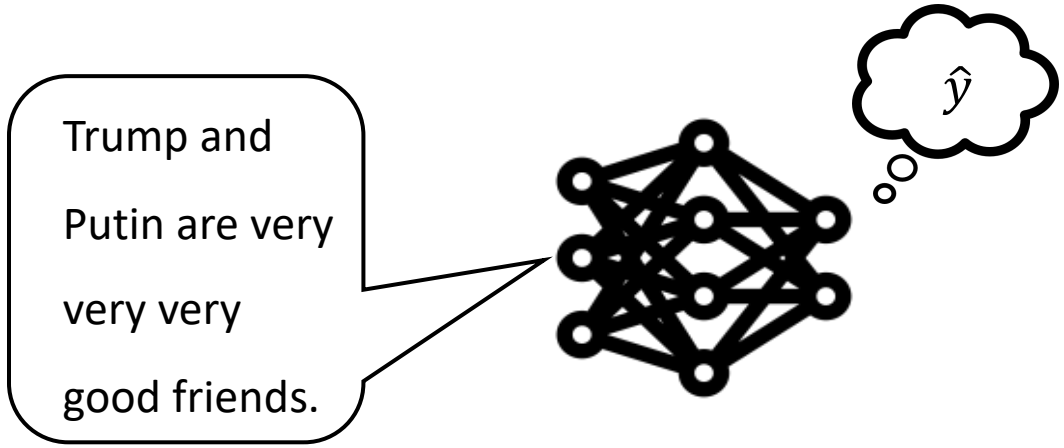
'Trump' = 1

'and' = 2

'Putin' = 3

…

?



Using a number to represent a word leads to following problem:

1. Close numbers means higher similarity, since their computational results are close.
2. Hard to justify which words should be assigned close values.

Word2Vec

In 2013, a team at Google led by Tomas Mikolov created word2vec.

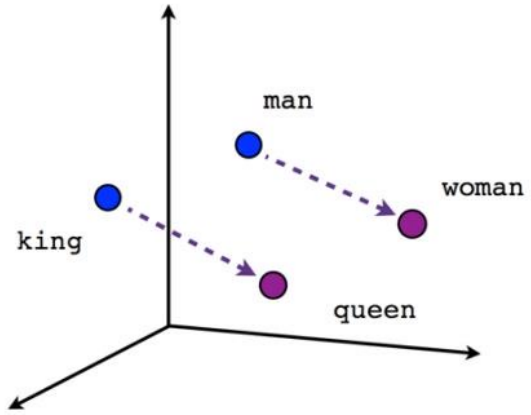
Stores each word with a vector of fixed number of dimensions (generally 200~300)

$$\text{e.g. } w^{Trump} = [0.44 \quad 0.56 \quad -0.7 \quad \dots \quad 0.81]$$

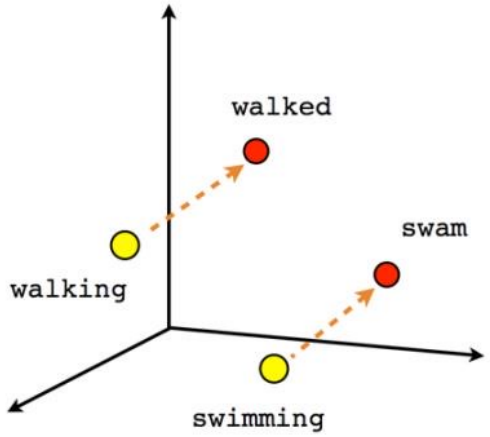
Word2Vec

Word2Vec model can capture Syntactical & Semantical relationship between words

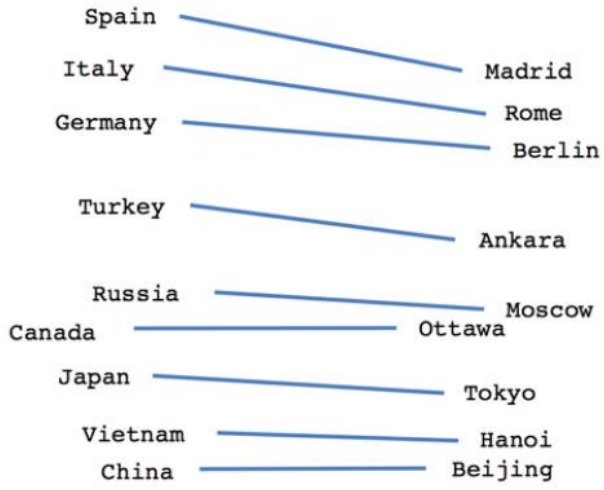
After reducing dimensions to 2-3 dimensions,



Male-Female

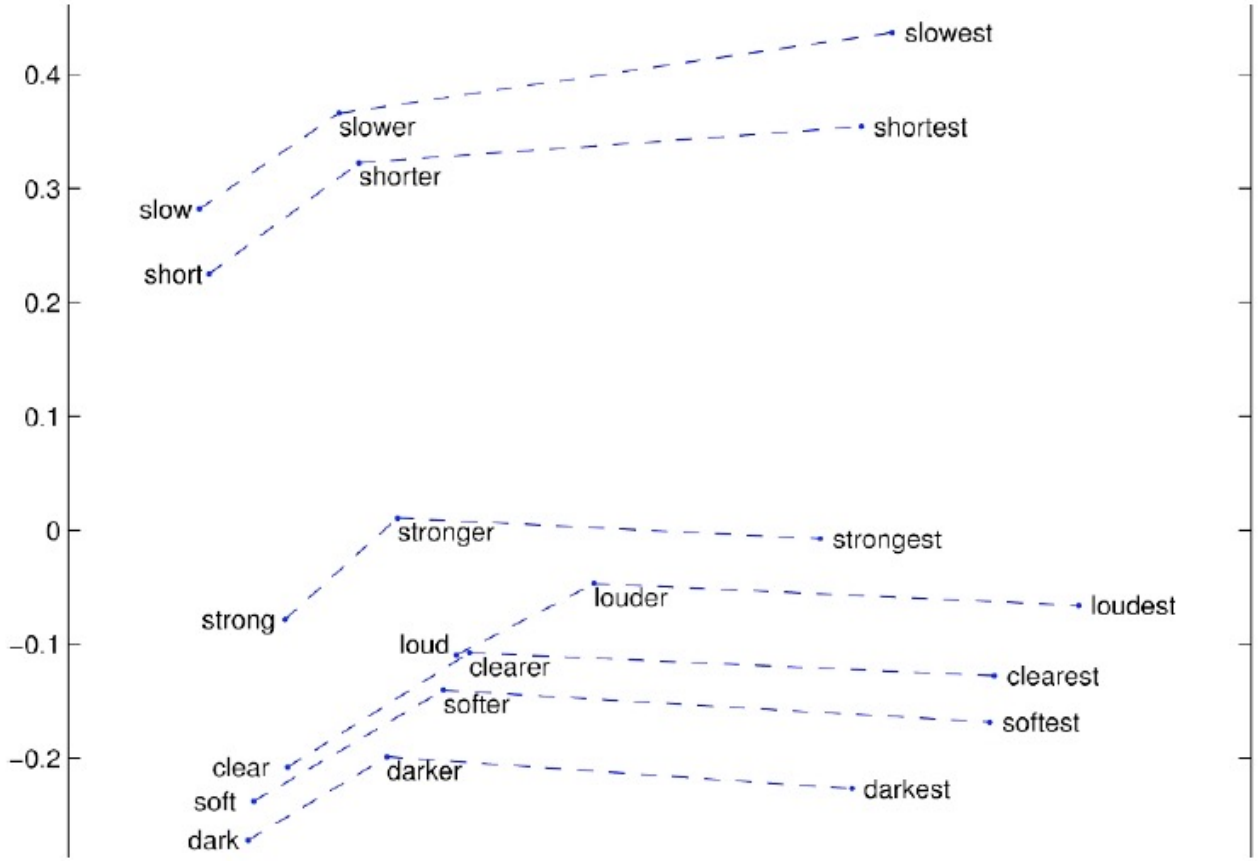


Verb tense



Country-Capital

Word2Vec



Primary-- Comparative--Superlative

Training Set

X

Trump	0.5	0.7	0.4	...	0.1
and	0.1	0.4	0.6	...	0.3

⋮

Word as a vector

Trump					
and					
Putin					
are					
very					
very					
very					
good					
friends					

Sentence as a matrix

Y

1
0
0

Positive

or

0
1
0

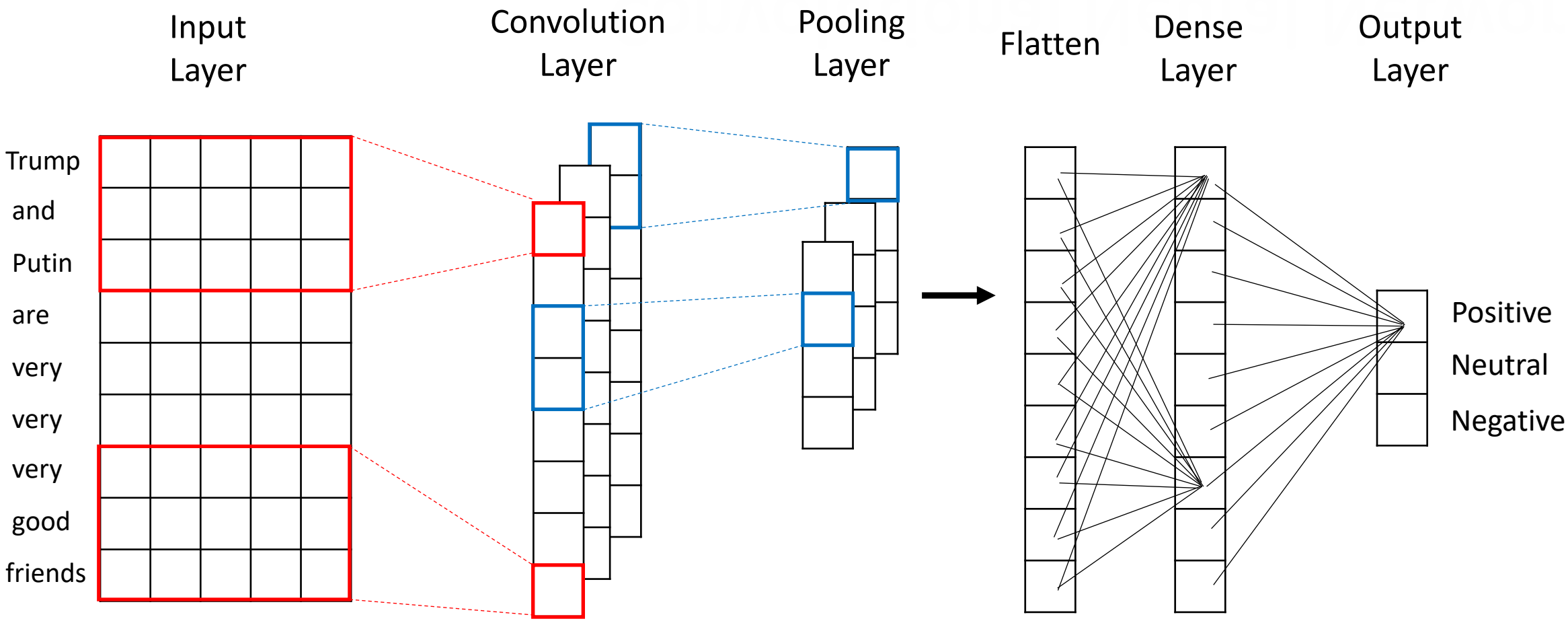
Neutral

or

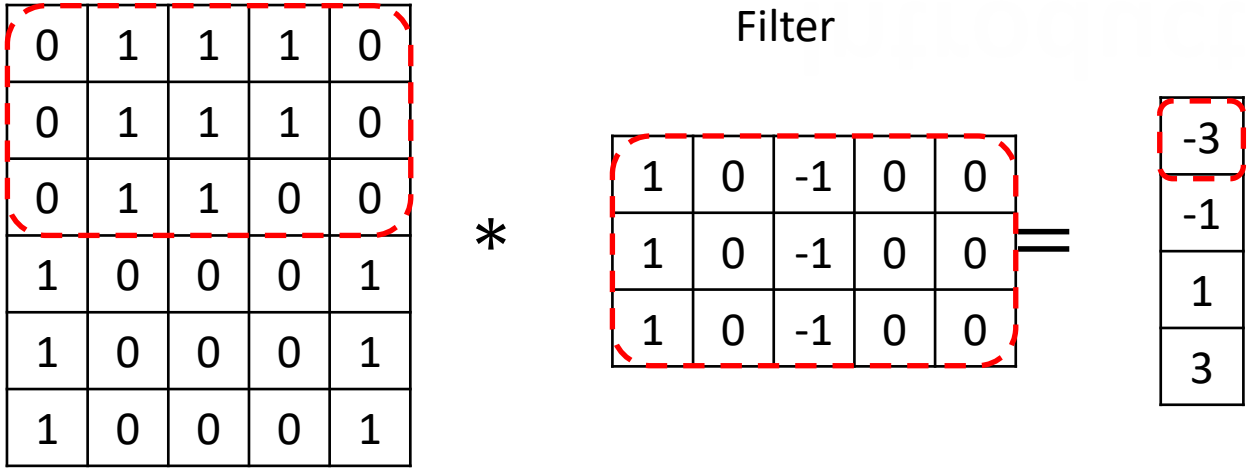
0
0
1

Negative

Convolutional Neural Network



Introduction: Convolution

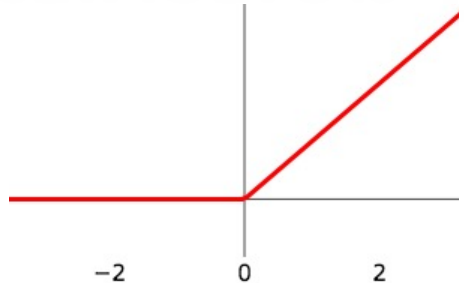


Convolution

Values in filter is parameters which need to be optimized

Introduction: Activation Function

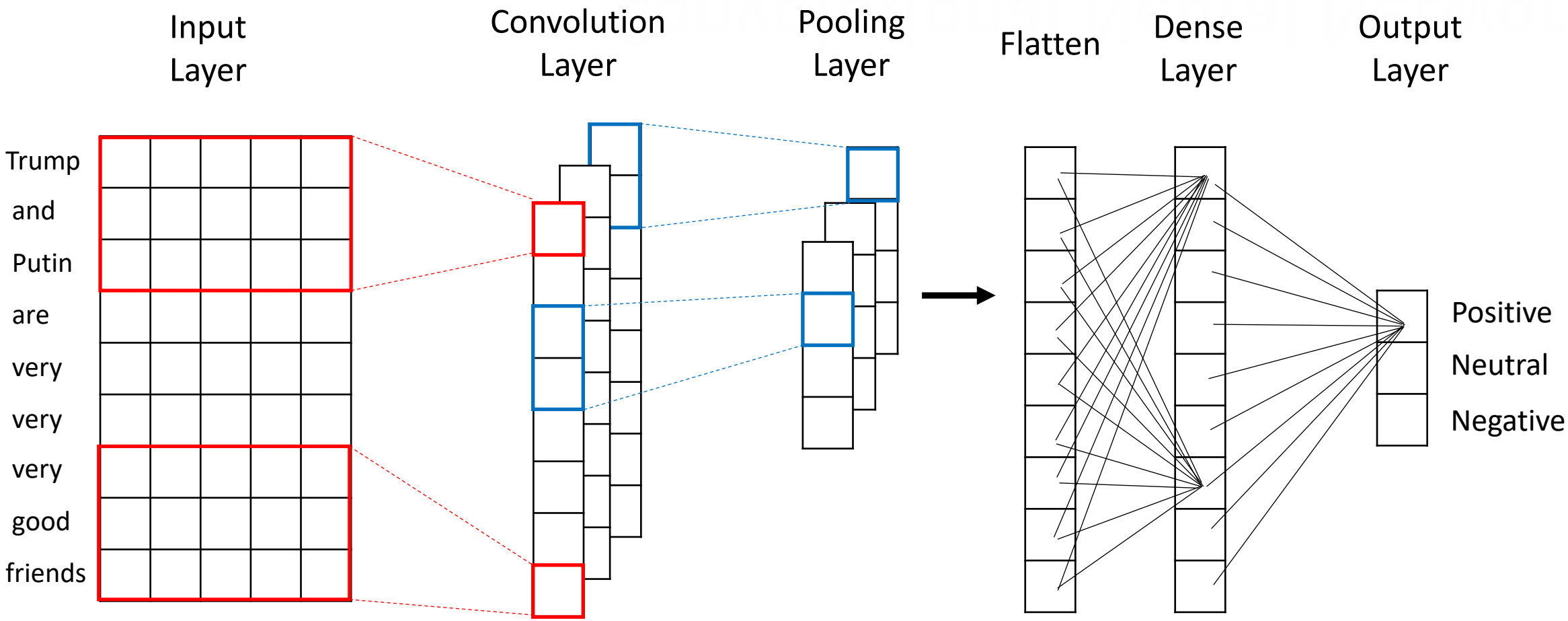
Relu function: $\mathrm{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$



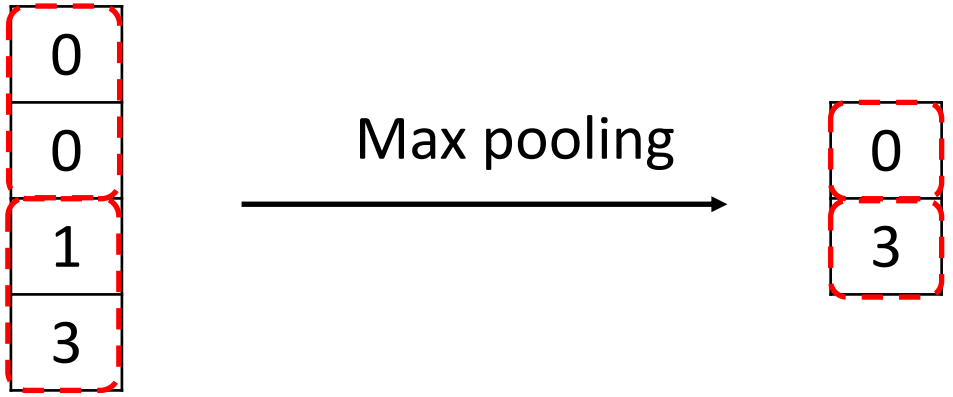
-3	ReLU(x) →	0
-1		0
1		1
3		3

Relu function can be replaced by other activation functions, e.g. Sigmoid function, Tanh function.

Convolutional Neural Network

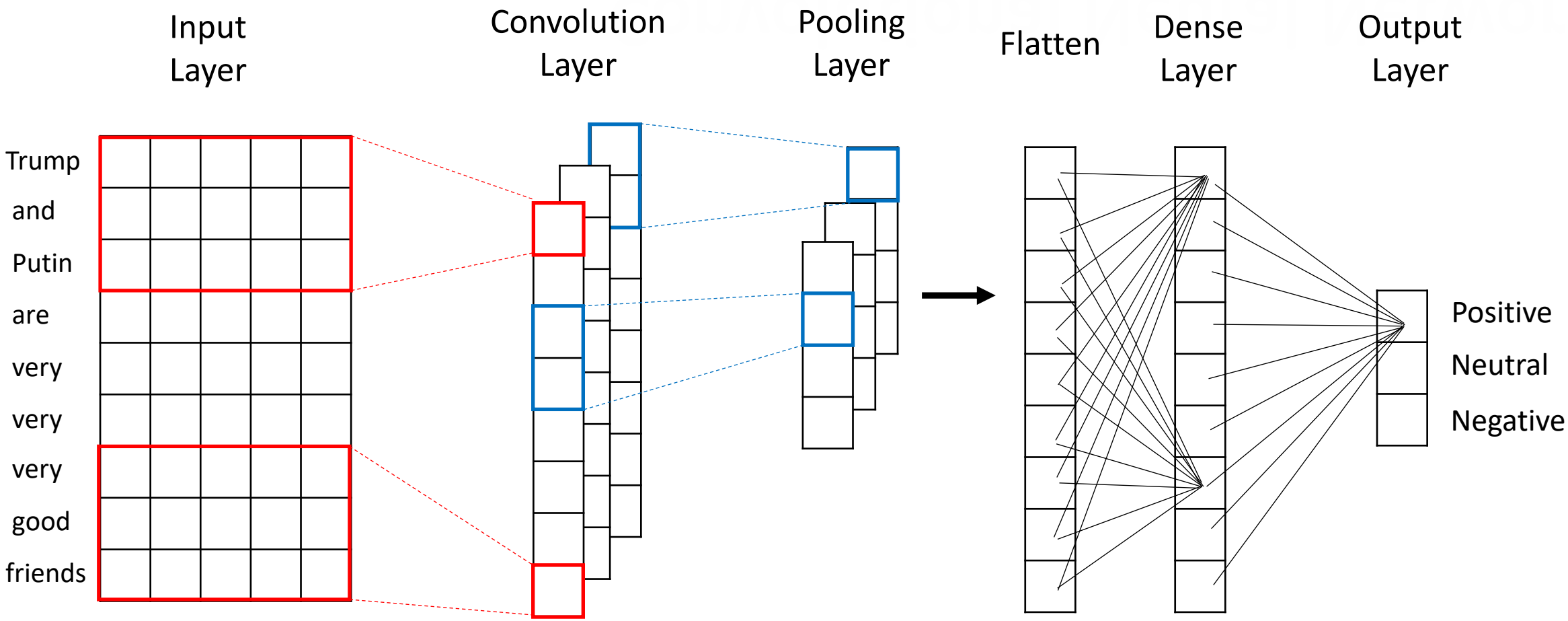


Introduction: Pooling

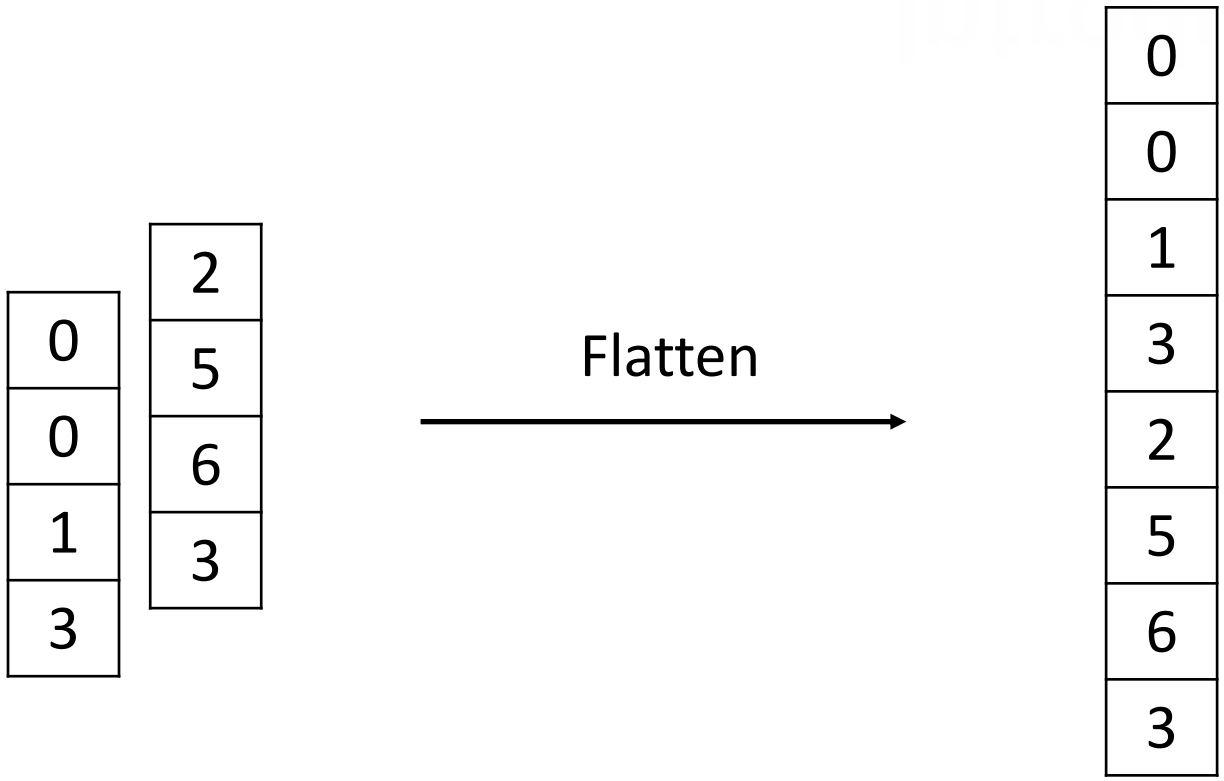


- There are other types of pooling, e.g. Average pooling.
- The function of pooling is to downsize data and accelerate the computation

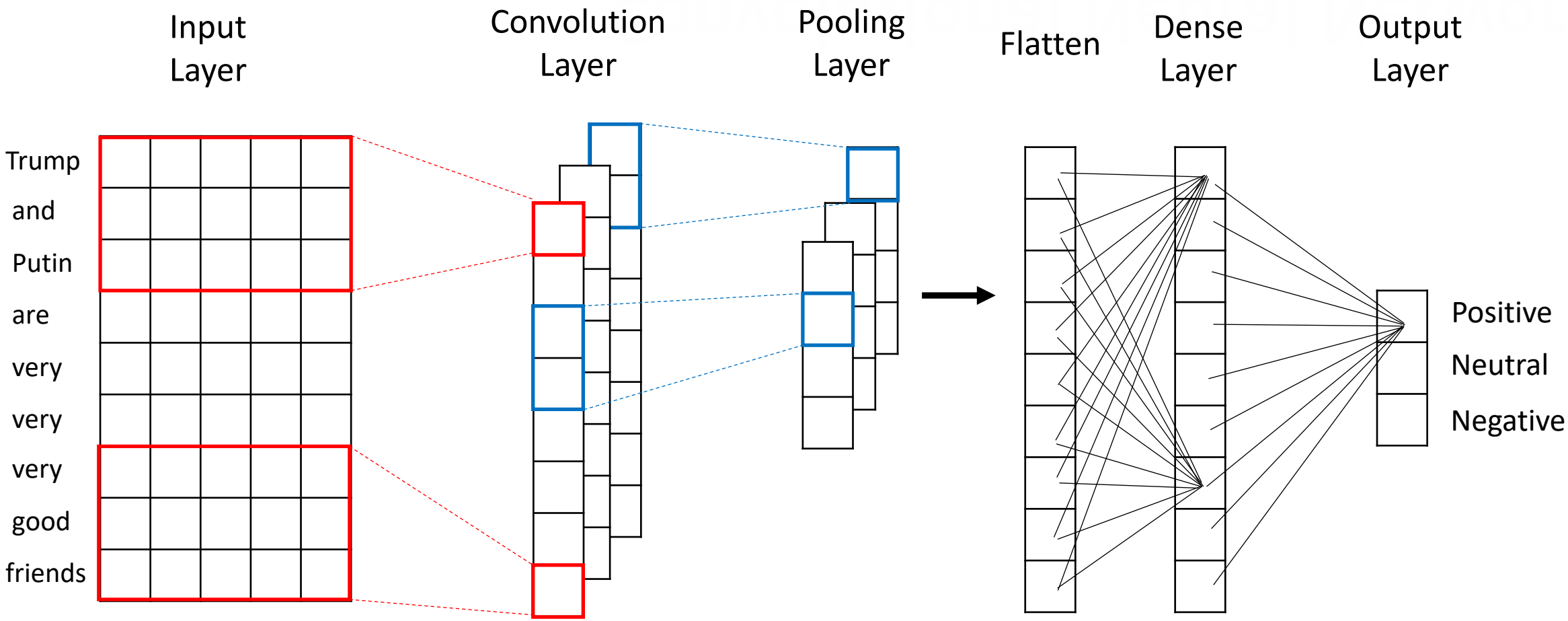
Convolutional Neural Network



Introduction: Flatten

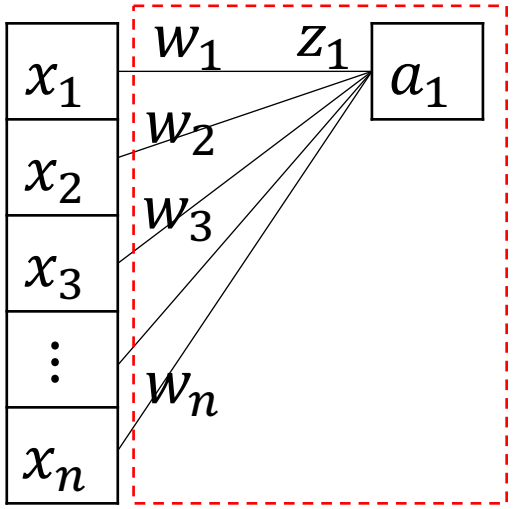


Convolutional Neural Network



Introduction: Dense Layer

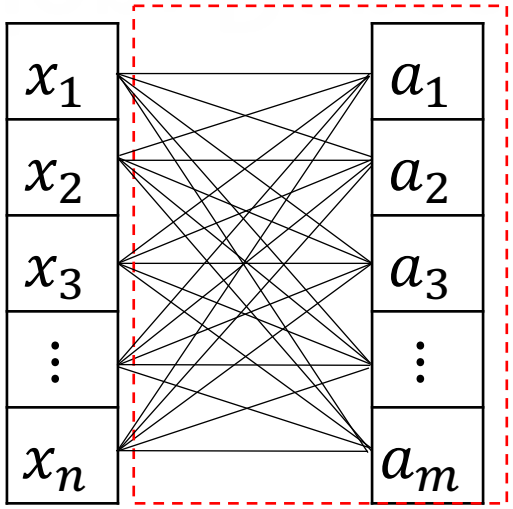
Perceptron



$$z_1 = \sum_{i=1}^{n} w_i x_i + b$$

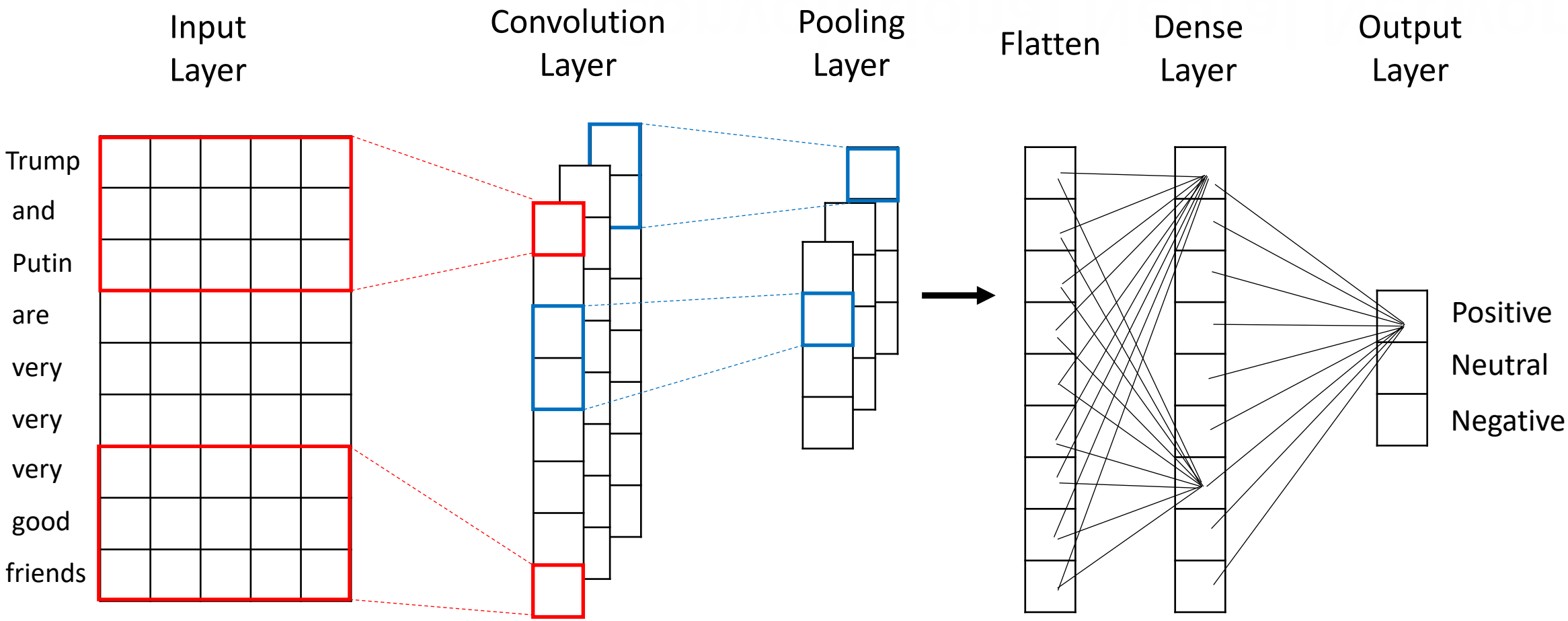
$$a_1 = Relu(z_1)$$

Dense Layer

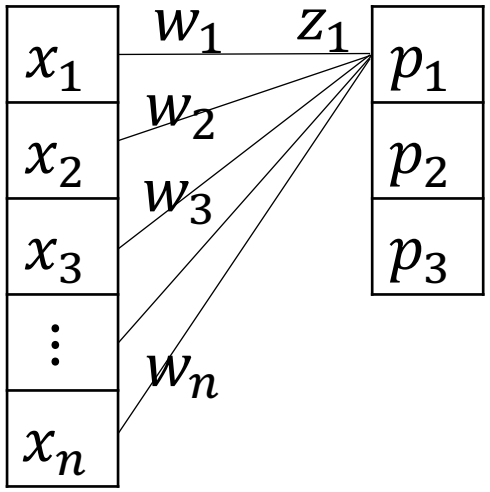


- w_i , b are the parameters which needs to be optimized.
- Relu function can be replaced by other activation functions, e.g. Sigmoid function, Tanh function.

Convolutional Neural Network



Introduction: Output Layer



Probability of 'Positive'

Probability of 'Neutral'

Probability of 'Negative'

Softmax function:

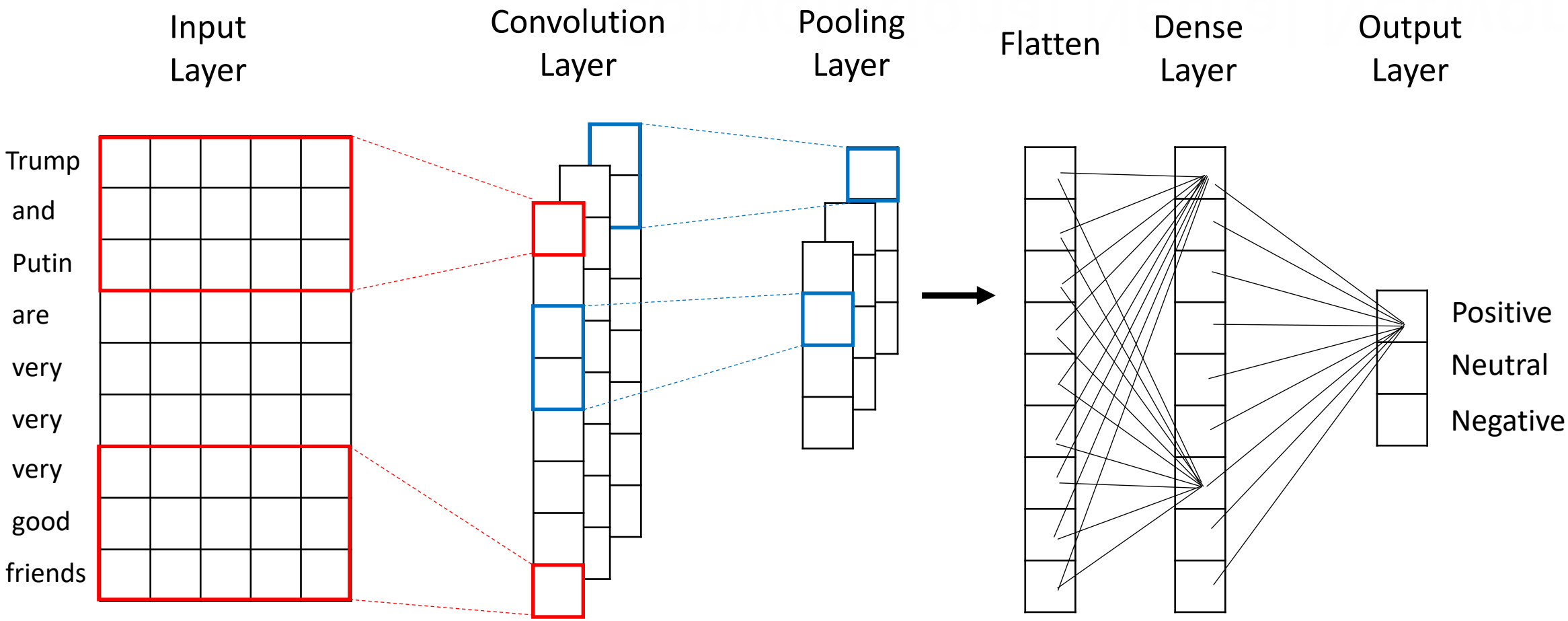
Generate the probability of each class

$$z_1 = \sum_{i=1}^{n} w_i x_i + b$$

$$p_1 = softmax(z_1) = \frac{e^{z_1}}{\sum_{k=1}^{3} e^{z_k}}$$

- w_i , b are the parameters which needs to be optimized.
- Softmax function can guarantee $p_1 + p_2 + p_3 = 1$.

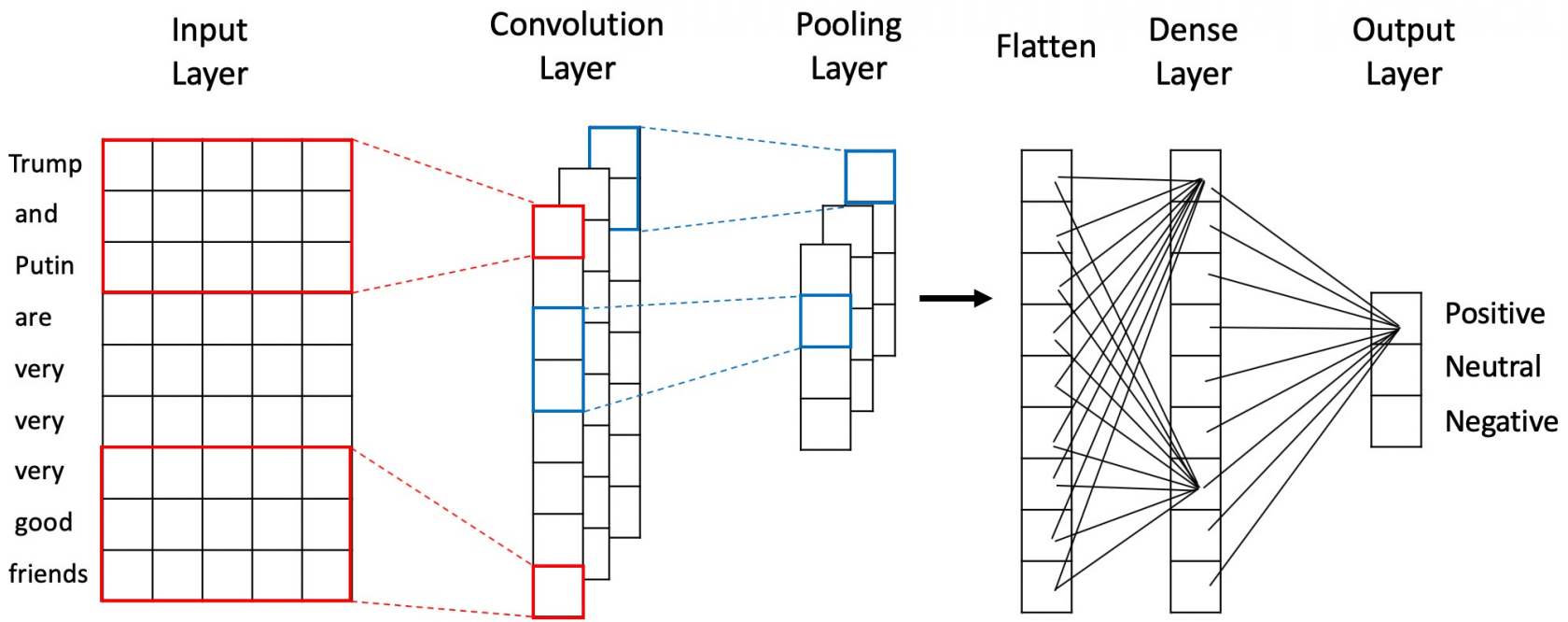
Convolutional Neural Network



How to train CNN

Loss function: Categorical Cross Entropy $$-\frac{1}{N}\sum_{i=1}^{N}\log p_{model}\left[y_i \in C_{y_i}\right]$$

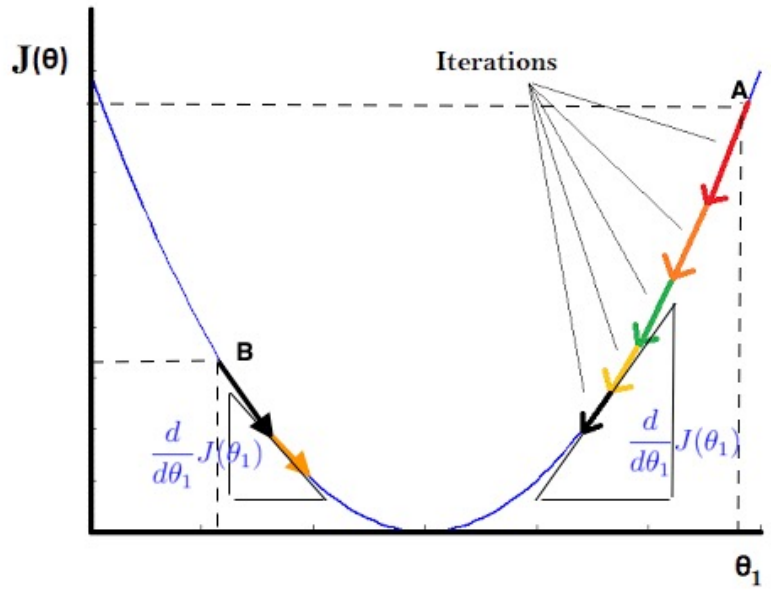
Calculate gradient for each parameter with Chain Rule $$\frac{dz}{dx}=\frac{dz}{dy}\cdot\frac{dy}{dx}$$

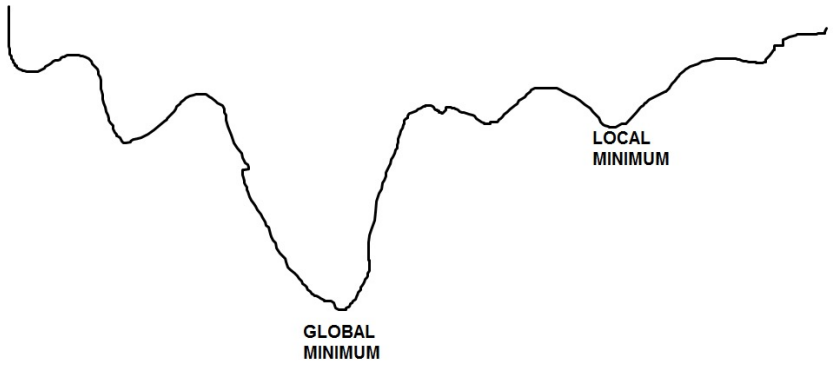


How to train CNN

Optimize parameters by repeating:

1. Forward Propagation: Calculate **y** with current parameter value
2. Backward Propagation: Update parameter value with Gradient Descent method.





Thank you for your kind attention!

Preprocessing Data

Basic Operations

1. Remove all URLs (e.g. www.xyz.com), hash tags (e.g. #topic), targets (@username)
2. Remove extra blank spaces, and line breaks
3. Remove all punctuations, symbols, numbers
4. Remove extra vowels repeated in sequence at least three times (e.g. cooooool and cool)
5. Expand acronyms (use an acronym dictionary)
6. Remove stopwords (e.g. the, is, at, which, and)
7. All text is converted to lower case

Emoticon Processing

Classify emoticons into two categories:

smile positive and smile negative

smile_positive	smile_negative
0:-)	>:(
:)	;(
:D	>:)
:*	D:<
:o	:(
:P	:|
;)	>:/

Angiani, Giulio, et al. "A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter." KDWeb. 2016.

Negation Processing

- All negative constructs (can’t, don’t, isn’t, never, etc.) are replaced with “not”.
- This procedure can enrich the corpus with more ‘not’.

Angiani, Giulio, et al. "A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter." KDWeb. 2016.

Stemming

- Put word variations into one bucket, e.g.

  {“great”, “greatly”, “greatest”, and “greater”} → “great”

- Stemming allows us to consider nouns, verbs and adverbs that have the same root in the same way.

Angiani, Giulio, et al. "A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter." KDWeb. 2016.

Correction of Spelling

- Detection and correction of misspelled words using a dictionary, e.g.
    1. substitute slang with its formal meaning (i.e., l8 → late).
    2. replace insults with the tag “bad word”.

Angiani, Giulio, et al. "A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter." KDWeb. 2016.