



We'll Be Starting Shortly!

To help us run the workshop smoothly, kindly:

- Submit all questions using the Q&A function
- If you have an urgent request, please use the "Raise Hand" function

Using Zoom: Viewing Mode





Side-By-Side Mode

- When sharing screen (slide share)
- With small thumbnails of people on the sidebar

STEPS:

- 1. View Options
- 2. Side-By-Side Mode





Sentiment Analysis With Tensorflow

Tom Huynh, 2021



Meet Tom Huynh

- Director of Machine Learning at Coderschool
- Former Head of Web Development at Weboptimizers, Melbourne
- Google Certified Tensorflow Developer for Deep Learning/Machine Learning
- Master of Al/Machine Learning, UNSW, Sydney
- Bachelor of Computer Science, University of Melbourne, Australia
- Winners and Runner-ups of multiple Startup/AI/Tech Hackathons





Who is Coderschool?

• CoderSchool is an organization that drives frontier technologies in Vietnam through education.



 Provide both online and offline coding courses, mentorship, and job placement for our students and alumni.



Kick-start Your Tech Journey With Us!

For Machine Learning lovers:

- Machine Learning & Data Science (online/offline and full-time/part-time)
- Data Analysis (part-time)



For aspiring Web Developers:

Full Stack Web Development (online/offline and full-time/part-time)

100% Job Guarantee for beginners with no coding experience:

For more information, please contact us!

Are you ready?

Email us at apply@coderschool.vn

Or call us at (+84) 984 326 569



Our client base



































































What I hope you take away from this

CODE LEAGUE 2021

- 1. What is Sentiment Analysis
- 2. Why it is very important to business and society
- 3. What is the general pipeline for preprocess text
- 4. How to represent words for computer to process (Tokenization)
- 5. What is Word Embedding Vector
- 6. Basic understanding of deep neural network
- 7. Build a deep learning model using Tensorflow (lab)



Let's think about a scenario



Text: "My beanie is amazing!..."

Length: 120 words





Text: "My shirt got wrong color!..."

Length: 300 words





Text: "The size of the shirt is really odd!..."

Length: 20 minutes





Text: "The new beanie feel really warm!..."

Length: 40 words







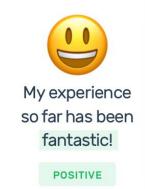
CoderSchool

Sentiment Analysis

1111111

The process of determining the opinion or feeling of a piece of text (positive, negative or neutral.)











Why is Sentiment Analysis important?

- Brand Reputation & Product/Service Monitoring
- Customer Service
- Market Research
- Social Purposes









Examples

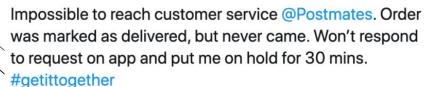
1111111













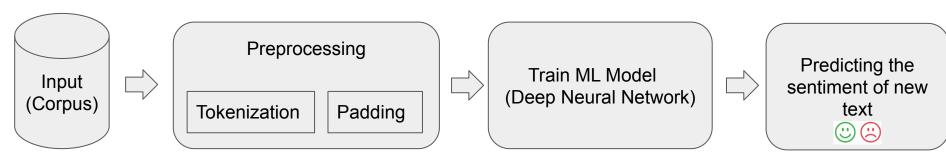
Replying to @AirbnbHelp

@airbnbhelp @airbnb I've had multiple conversations with your customer support team and they are absolutely worthless. No one has been able to address my issue or process the refund I was promised.



The process of sentiment analysis







IMDB Movie Review Dataset

label

text



Text (corpus):

25,000 movie reviews from IMDB website

Label (expected output):

0 means negative sentiment1 means positive sentiment

Split Training/Test:

25,000 for training 25,000 for testing

0	This was an absolutely terrible movie. Don't be lured in by Christopher Walken or Michael Ironside.
(neg)	Both are great actors, but this must simply be their worst role in history. Even their great acting coul
	not redeem this movie's ridiculous storyline. This movie is an early nineties US propaganda piece.
	The most pathetic scenes were those when the Columbian rebels were making their cases for
	The most pathetic scenes were those when the Columbian rebels were making their cases for

The most pathetic scenes were those when the Columbian rebels were making their cases for revolutions. Maria Conchita Alonso appeared phony, and her pseudo-love affair with Walken was nothing but a pathetic emotional plug in a movie that was devoid of any real meaning. I am disappointed that there are movies like this, ruining actor's like Christopher Walken's good name. I could barely sit through it.

This is the kind of film for a snowy Sunday afternoon when the rest of the world can go ahead with (pos) its own business as you descend into a big arm-chair and mellow for a couple of hours. Wonderful performances from Cher and Nicolas Cage (as always) gently row the plot along. There are no rapids to cross, no dangerous waters, just a warm and witty paddle through New York life at its best. A family film in every sense and one that deserves the praise it received.







How do we turn sentences into number? (Tokenization)



How can a text be understood by a computer?



There are two main ways to tokenize a text:

- Each letter
- Each word



Tokenize a letter



 Let's start with a very naive way to tokenize a letter to be understood by a computer

A	В	C	D	E	F	G	H	I	J	K	L	M
1	1	1	1	1	1	1	1	1	1	1	1	1
65	66	67	68	69	70	71	72	73	74	75	76	77
N	0	Р	Q	R	S	Т	U	V	W	Х	Y	Z
1	1	1	1	1	1	1	1	1	1	1	1	1
78	79	80	81	82	83	84	85	86	87	88	89	90

Table 1: ASCII encodings of English capital letters.





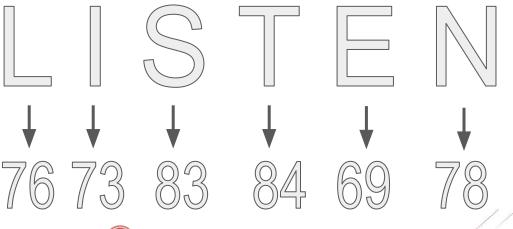
Example tokenize letter using ASCII

1111111

A	В	C	D	Ε	F	G	Н	I	J	K	L	M
1	1	1	1	1	1	1	1	1	1	1	1	1
65	66	67	68	69	70	71	72	73	74	75	76	77
N	0	P	Q	R	S	Т	U	V	W	Х	Y	Z
1	1	1	1	1	1	1	1	1	1	1	1	1
78	79	80	81	82	83	84	85	86	87	88	89	90



Table 1: ASCII encodings of English capital letters.





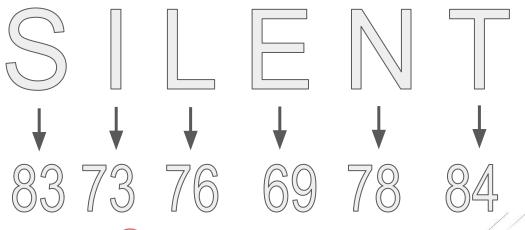
Example tokenize letter using ASCII

1111111

Α	В	C	D	Ε	F	G	Н	I	J	K	L	M
1	1	1	1	1	1	1	1	1	1	1	1	1
65	66	67	68	69	70	71	72	73	74	75	76	77
N	0	P	Q	R	S	Т	U	V	W	Х	Y	Z
1	1	1	1	1	1	1	1	1	1	1	1	1
78	79	80	81	82	83	84	85	86	87	88	89	90



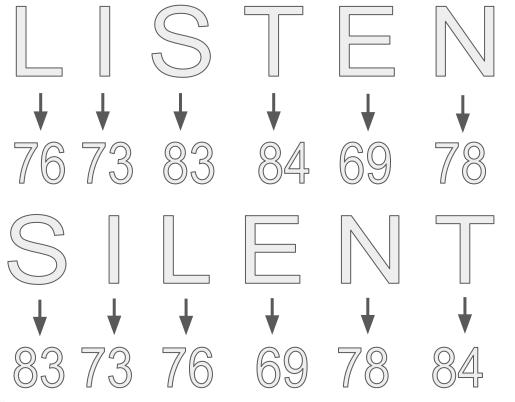
Table 1: ASCII encodings of English capital letters.





Problems of Letter Tokenization

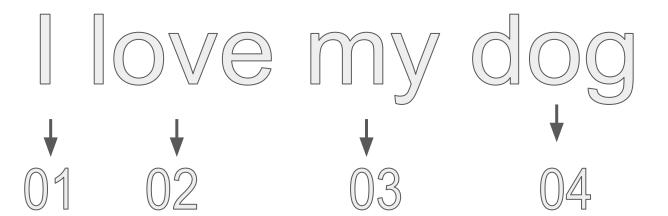
1111111





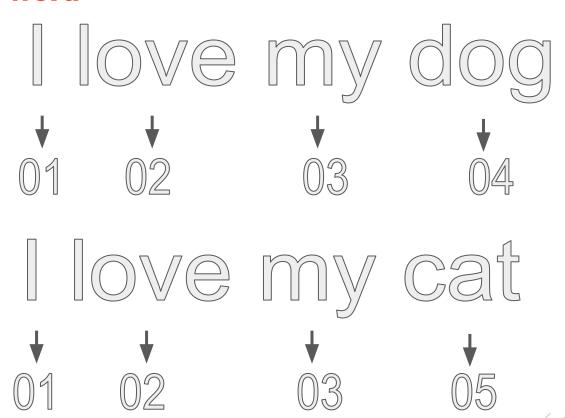
Tokenize a word







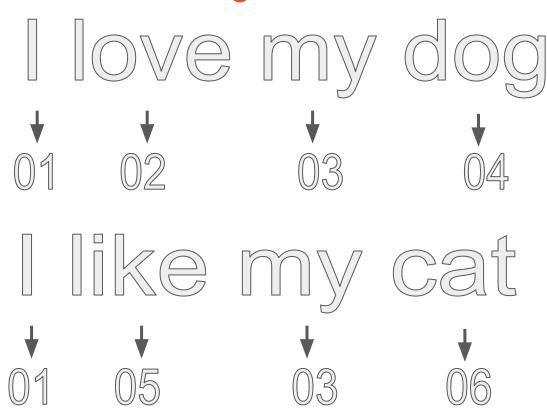
Tokenize a word







Tokenize a word - not enough? Problem?



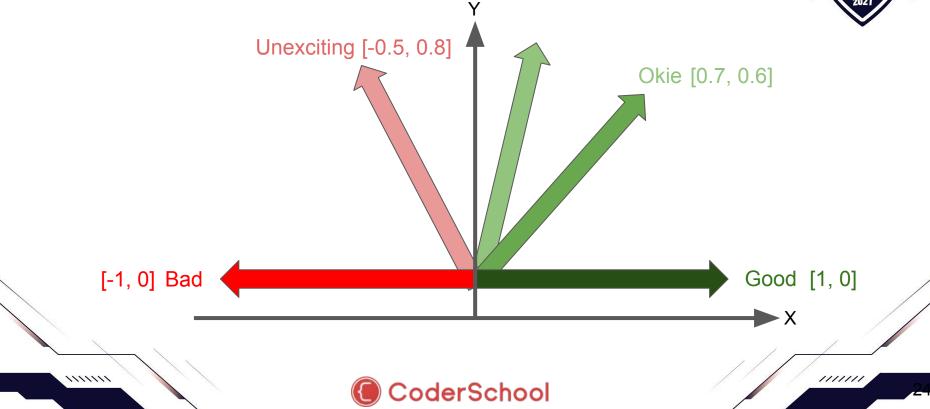




Word Embedding Vector

"Okie but Unexciting!" - Tom





Simple Word Embedding Vectors

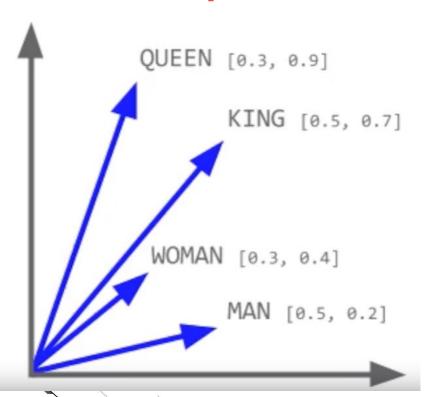


93	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	-1	1	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	0.93	0.95	-0.01	0.00
Age	0.03	0.02	0.70	0.69	0.03	-0.02
Food	0.09	0.01	0.02	0.01	0.95	0.97



Relationships from Embedding Vectors





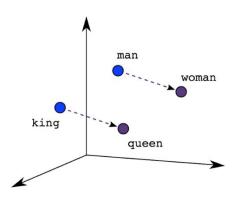
1111111

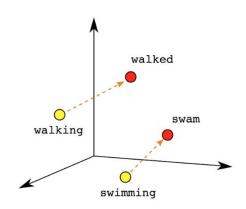
What is king + man - woman?

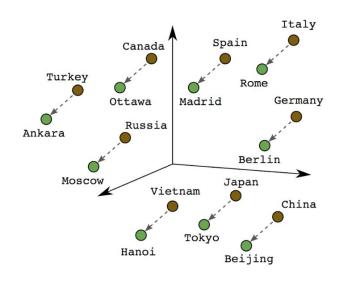


Relationships in Word Embedding









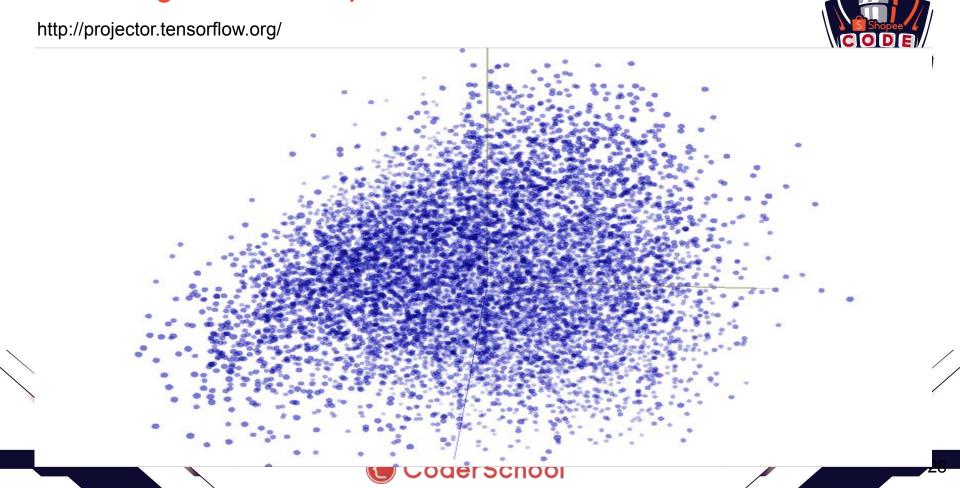
Male-Female

Verb Tense

Country-Capital



Embedding Vectors of 10,000 words for 200 dimensions



Word Embedding Space in Sentiment Analysis



- wonderful
- wonderfully

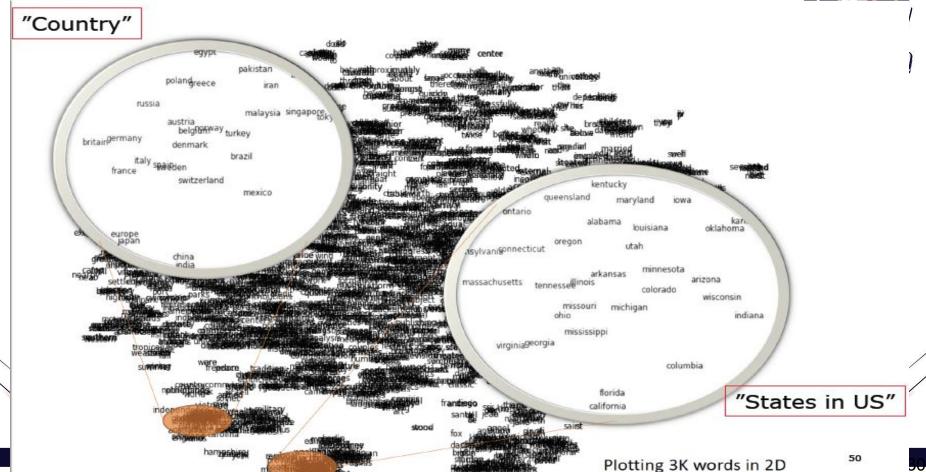
1111111

horrible



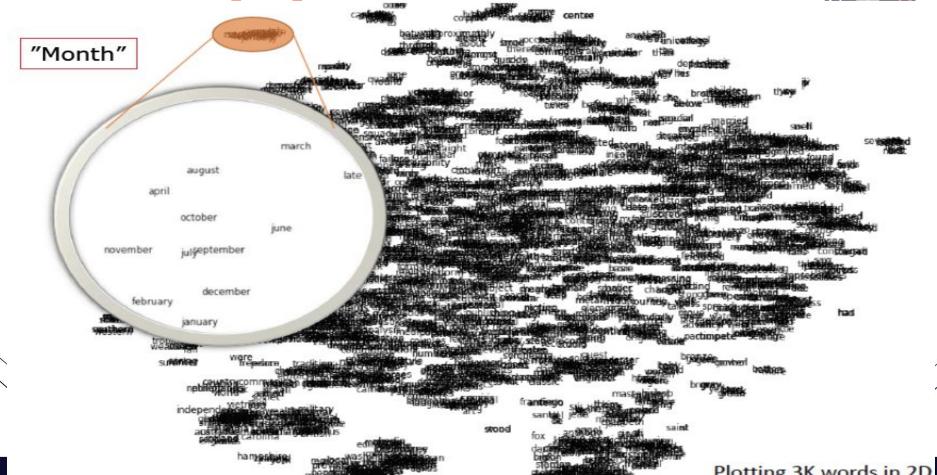
Word Embedding in general data





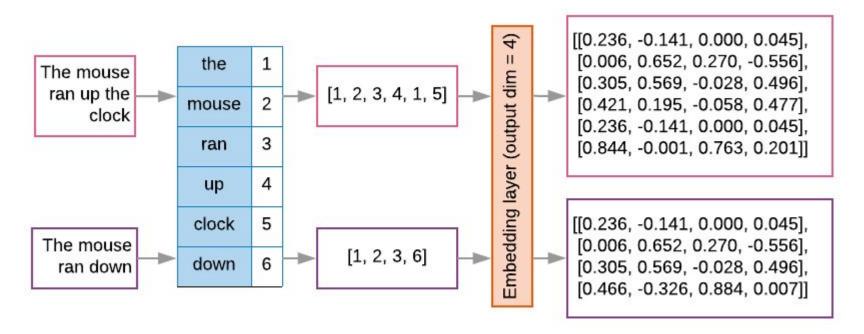
Word Embedding in general data





The Pipeline of Tokenization and Embedding Layer

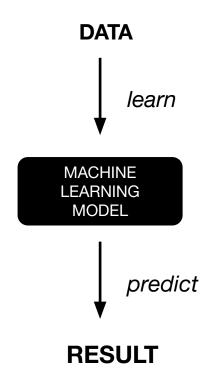






Machine Learning Model

1111111





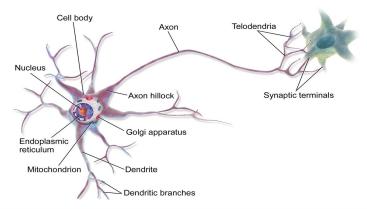
The Rise of Deep Neural Network since 2010

1111111





Deep Learning



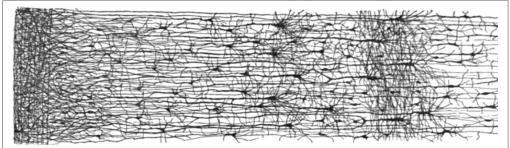


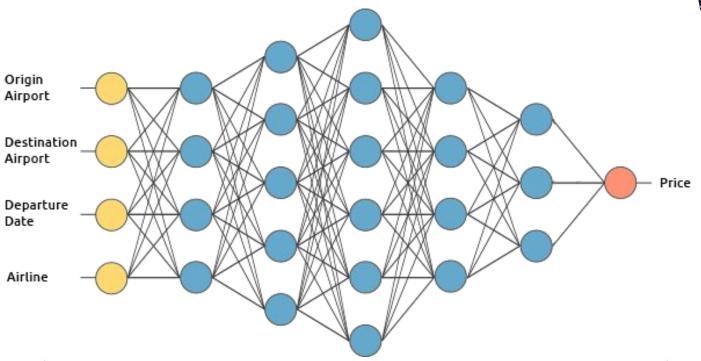
Figure 10-2. Multiple layers in a biological neural network (human cortex)⁶





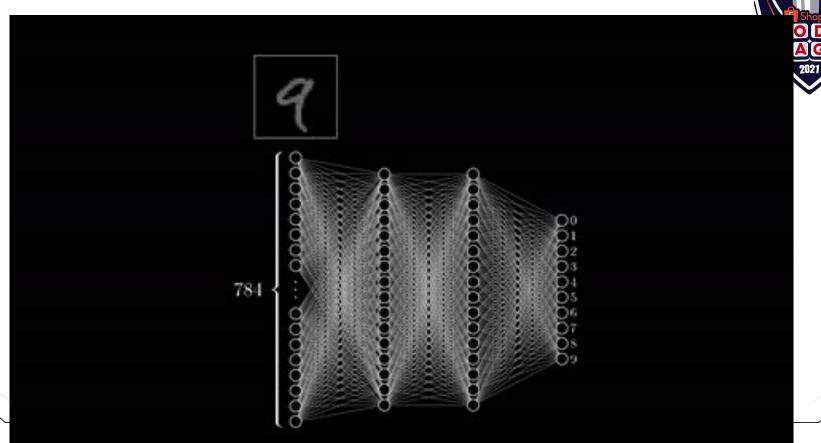
An Simple Example of Neural Network

1111111





Neural Network in Action



CoderSchool







Hand-on Lab Link

1111111

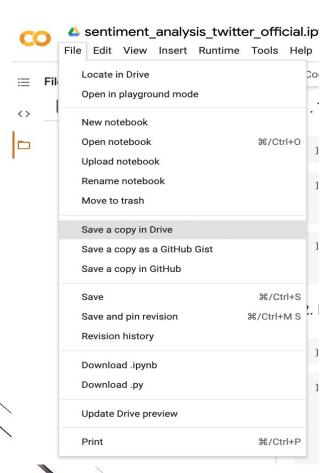




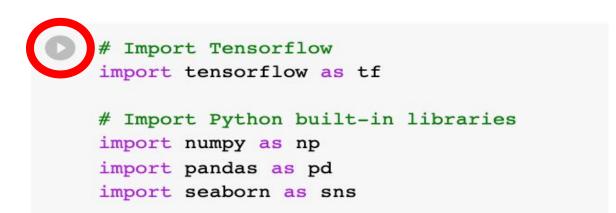
https://bit.ly/3ezu6IP







1111111



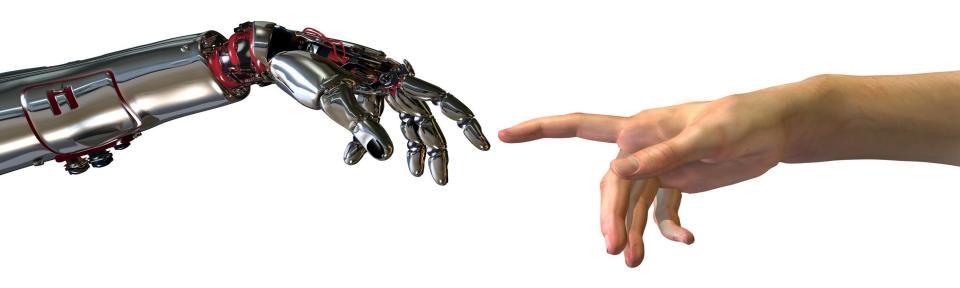
or Shift + Enter

Runtime — Change Runtime Type
Hardware Accelerator — GPU



Thank you for your listening









Extra Datasets to play around



- Twitter airline sentiment on Kaggle: https://www.kaggle.com/crowdflower/twitter-airline-sentiment
- Amazon Product Reviews for Sentiment Analysis: https://www.kaggle.com/bittlingmayer/amazonreviews
- Product Reviews: https://www.kaggle.com/bittlingmayer/amazonreviews
- IMDB Movie Reviews:
 https://www.kaggle.com/lakshmi25npathi/sentiment-analysis-of-imdb-movie-reviews
- Amazon Fine Food Reviews:
 https://www.kaggle.com/snap/amazon-fine-food-reviews



Extra Materials to learn NLP on Tensorflow



- Word Embedding: https://www.tensorflow.org/tutorials/text/word-embeddings
- Word2Vec: https://www.tensorflow.org/tutorials/text/word2vec
- Text Classification with RNN:
 https://www.tensorflow.org/tutorials/text/text_classification_rnn



Your Feedback Matters!





