

Introduction To Deep Learning

Hyunbeen Kim

been000904@gmail.com

NLP Team

2024/03/12

Contents

- NLP Team Rule
- Self Test
- Basic ML (Linear Regression)
- Introduction To Deep Learning
- Introduction To NLP
- Tip for Study

NLP Team Rule

1. 매주 화요일 20시~22시 사당역 Moim에서 진행
2. 커리큘럼은 주제와 관련된 논문 리뷰하는 방식으로 진행
3. 한두명의 발제자가 한 주제와 관련한 논문 3편 정도 리뷰
4. 발제자가 아니더라도 웬만하면 다음 진행할 논문은 다 읽어오기 !
5. 디스커션시 과대망상 환영 !
6. 프로젝트는 자유롭게 진행 ! 함께 하고 싶으면 팀에 제안하는 식으로

Self Test

모 스타트업의 지원 요건

▼ 딥러닝 질문: 10개 (삼각형을 클릭하세요)

- Gradient Descent란?
- Sigmoid의 단점은?
- Validation 세트, Test 세트의 각각의 역할은?
- Auto Encoder란?
- Dropout의 효과는?
- CNN의 장점은?
- Word2Vec의 원리는?
- Adam Optimizer의 동작은?
- Batch Normalization의 동작은?
- CycleGAN이란?

▼ 딥러닝 모델 개발자 기술 질문: 9개 중 5개 이상 잘 대답할 수 있어야 함 (삼각형을 클릭하세요)

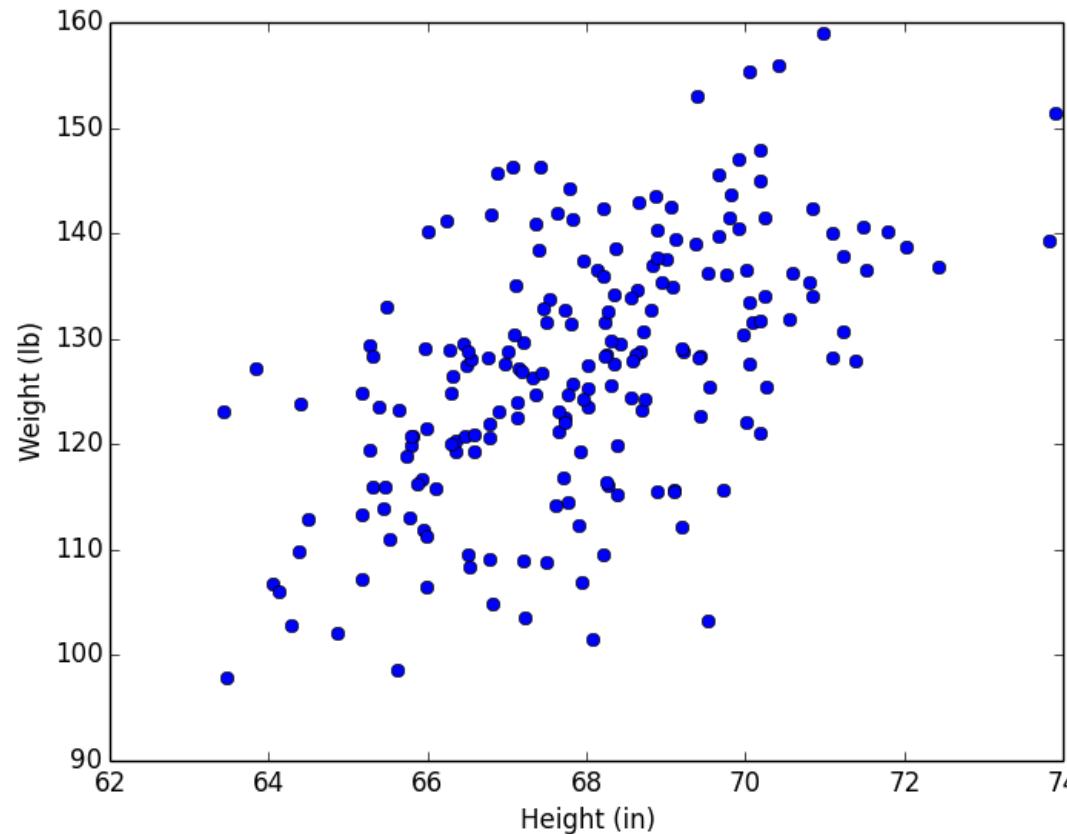
- Gradient Descent란?
- Loss Surface란?
- Attention이란?
- Transformer란?
- Collaborative filtering이란?
- Few-Shot Learning이란?
- Federated Learning이란?
- SVD란?
- 중심극한정리란?

Self Test

What is ML/DL?

Basic Machine Learning (Linear Regression)

Weight Prediction

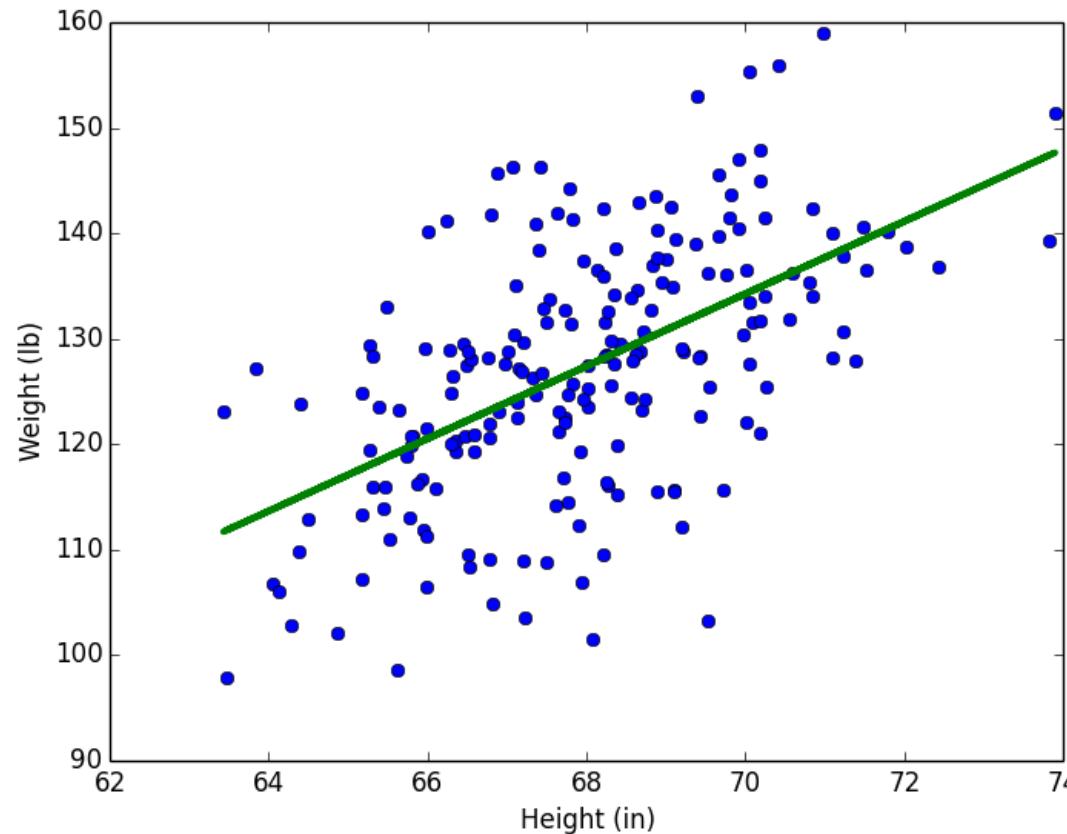


Q1.

몸무게를 맞추고자 할 때 생각할 수 있는
좋은 방법은 무엇일까?

Basic Machine Learning (Linear Regression)

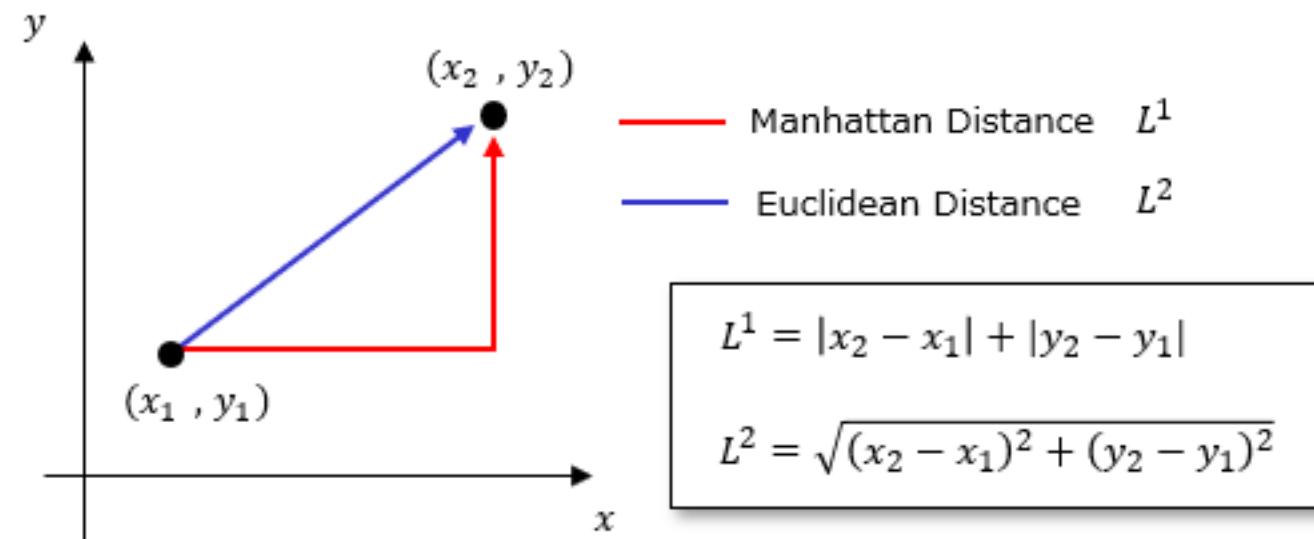
Weight Prediction



Q2.
초록선을 구하기 위해서 어떤 방식을 취할
수 있을까?

Basic Machine Learning (Linear Regression)

Distance Metrics



Basic Machine Learning (Linear Regression)

Method For Linear Regression

Normal Equation

$$A^T A \hat{x} = A^T b \rightarrow \hat{x} = (A^T A)^{-1} A^T b$$

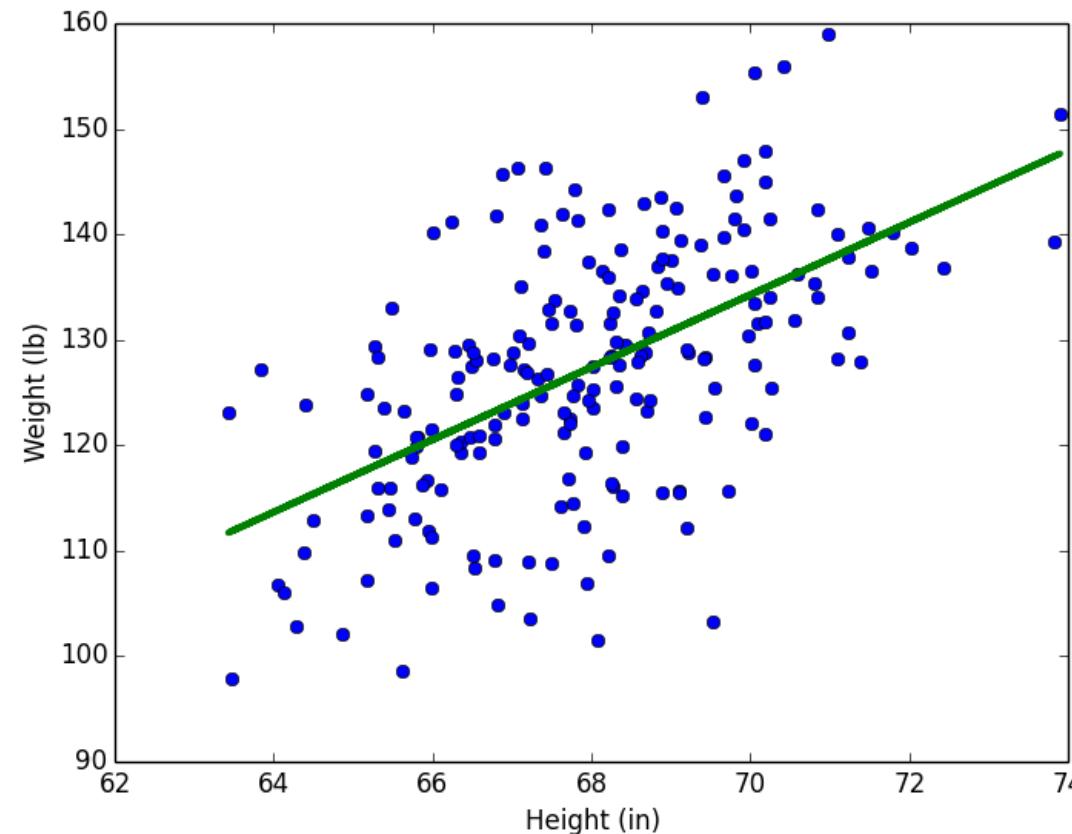
Coefficient

$$y = \text{intercept} + \text{coefficient} \times x_{value}$$

$$\text{coefficient} = \beta_1 = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

Basic Machine Learning (Linear Regression)

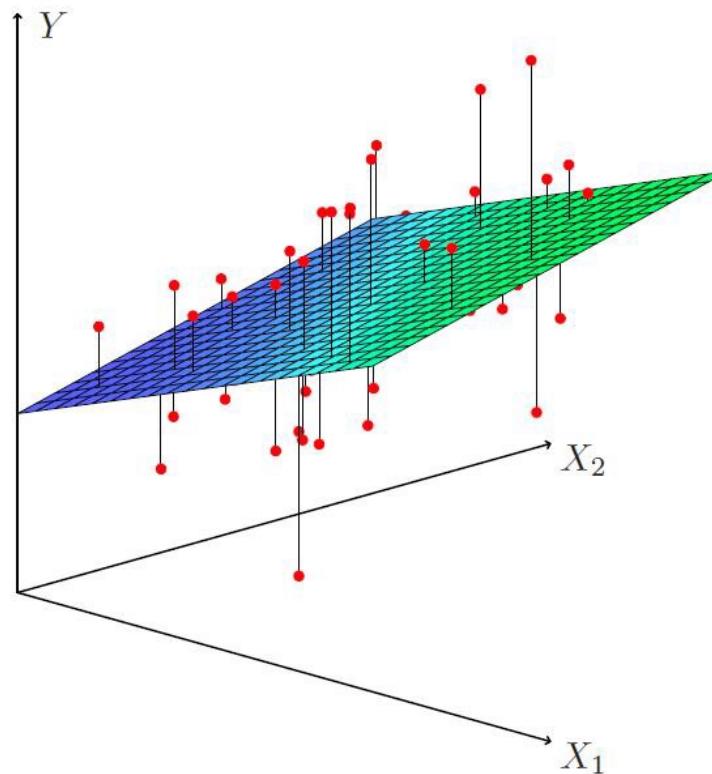
Weight Prediction



Q3.
몸무게 예측을 위해 키 피쳐만을 사용하는
것이 최선일까?

Basic Machine Learning (Linear Regression)

Weight Prediction with Multi Variable



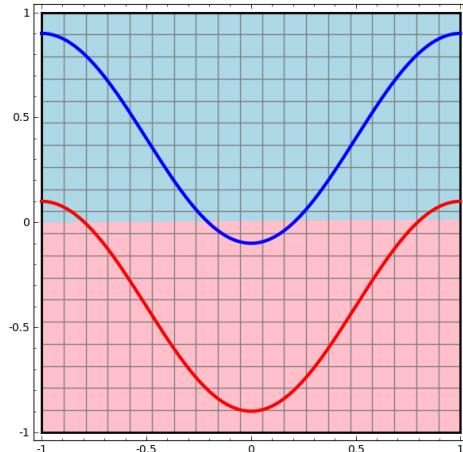
Input:
Scala -> Vector

Output:
Scala -> Scala

Space:
2D -> 3D

Basic Machine Learning (Linear Regression)

Limit of Linear Regression



Non-Linear Pattern

Q4.
직선 하나로 두 곡선을
분류할 수 있을까?



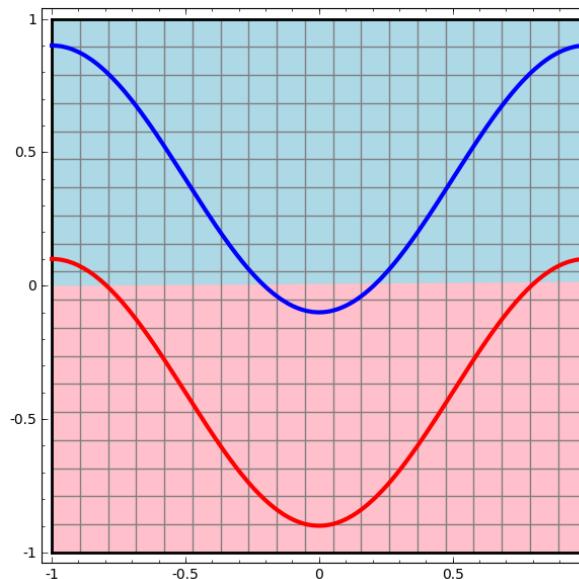
Q5.
이미지 가진 숫자들을 의미 있는
벡터로 볼 수 있을까?



Unstructured Data

Introduction To Deep Learning

How to Solve Non-Linear Problem



Q6.

두 곡선을 분류할 수 있도록 변환을 가하면
되지 않을까? 근데 어떻게?

Introduction To Deep Learning

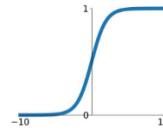
How to Solve Non-Linear Problem

Linear Transformation With Non Linear Function

Activation Functions

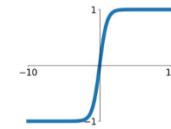
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



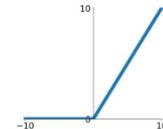
tanh

$$\tanh(x)$$



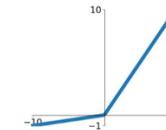
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

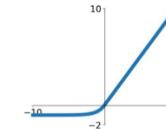


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Q7.

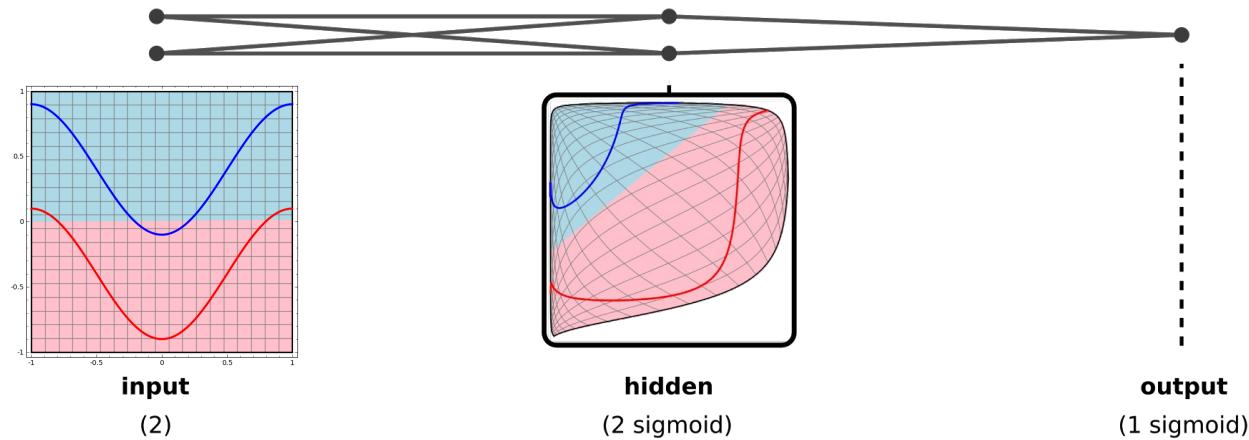
Activation Function 연산은 가역일까?

Q8.

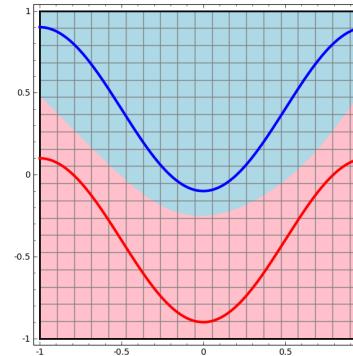
Activation Function 연산은 데이터가 놓인
Space의 Rank를 바꿀 수 있을까?

Introduction To Deep Learning

How to Solve Non-Linear Problem



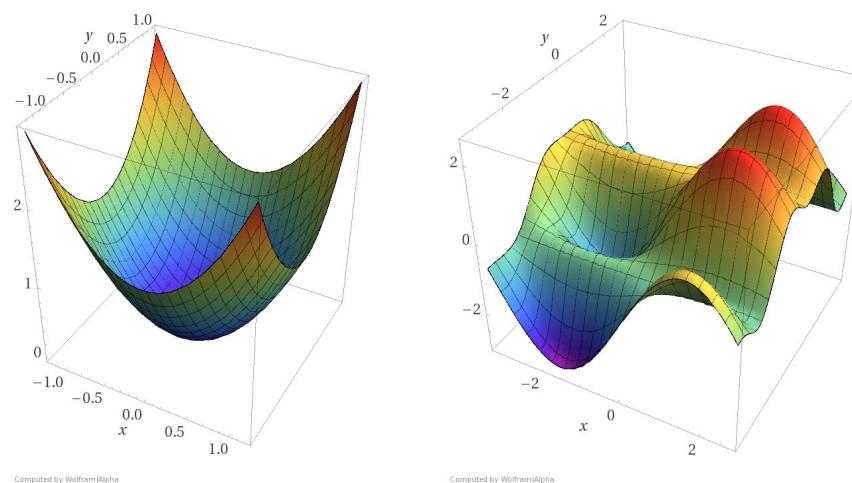
Result:



Introduction To Deep Learning

Non-Linear Optimization

Linear Regression Have Convex Function as Loss Function,
But Non-Linear Regression Have Non-Convex Function as Loss Function
We can't compute analytic solution, So we must compute numerical solution



Introduction To Deep Learning

Non-Linear Optimization

Numerical Solution: Gradient Descent

Cost Function

$$J(\Theta_0, \Theta_1) = \frac{1}{2m} \sum_{i=1}^m [h_\Theta(x_i) - y_i]^2$$

↑
Predicted Value ↑ True Value

Gradient Descent

$$\Theta_j = \Theta_j - \alpha \frac{\partial}{\partial \Theta_j} J(\Theta_0, \Theta_1)$$

↑
Learning Rate

Now,

$$\begin{aligned}\frac{\partial}{\partial \Theta} J_\Theta &= \frac{\partial}{\partial \Theta} \frac{1}{2m} \sum_{i=1}^m [h_\Theta(x_i) - y_i]^2 \\ &= \frac{1}{m} \sum_{i=1}^m (h_\Theta(x_i) - y) \frac{\partial}{\partial \Theta_j} (\Theta x_i - y) \\ &= \frac{1}{m} (h_\Theta(x_i) - y) x_i\end{aligned}$$

Therefore,

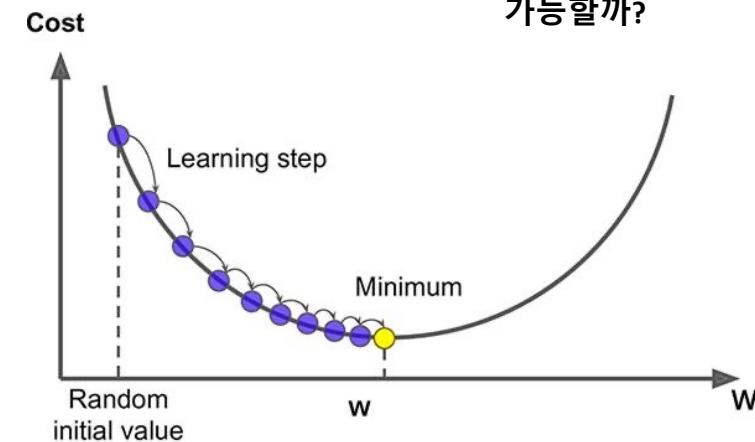
$$\Theta_j := \Theta_j - \frac{\alpha}{m} \sum_{i=1}^m [(h_\Theta(x_i) - y) x_i]$$

Q9.

Learning Rate는 무슨 의미를 가질까?

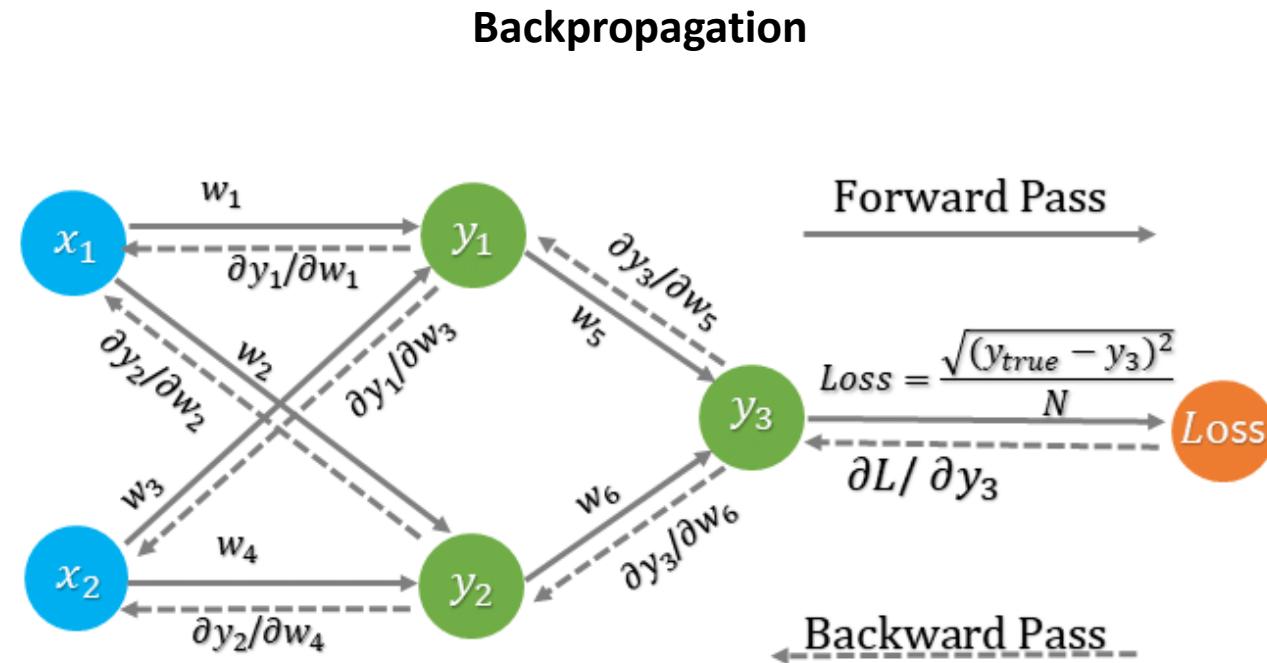
Q10.

합성함수에 대해서도 Gradient Descent가 가능할까?



Introduction To Deep Learning

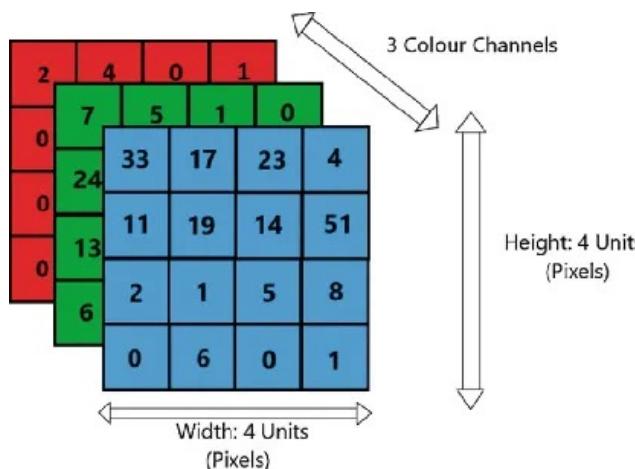
Non-Linear Optimization



Introduction To Deep Learning

How to Model Unstructured Data

Natural Language and Image are byte data in computer

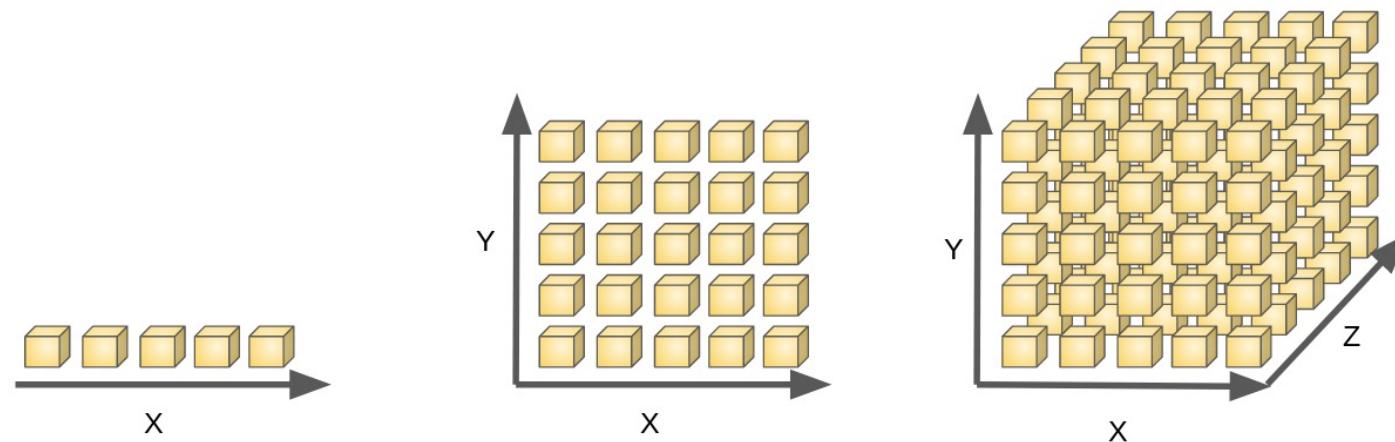


[(base) been@Been-MacBook-Pro math % ascii -d]									
0 NUL	16 DLE	32	48 0	64 @	80 P	96 `	112 p		
1 SOH	17 DC1	33 !	49 1	65 A	81 Q	97 a	113 q		
2 STX	18 DC2	34 "	50 2	66 B	82 R	98 b	114 r		
3 ETX	19 DC3	35 #	51 3	67 C	83 S	99 c	115 s		
4 EOT	20 DC4	36 \$	52 4	68 D	84 T	100 d	116 t		
5 ENQ	21 NAK	37 %	53 5	69 E	85 U	101 e	117 u		
6 ACK	22 SYN	38 &	54 6	70 F	86 V	102 f	118 v		
7 BEL	23 ETB	39 '	55 7	71 G	87 W	103 g	119 w		
8 BS	24 CAN	40 (56 8	72 H	88 X	104 h	120 x		
9 HT	25 EM	41)	57 9	73 I	89 Y	105 i	121 y		
10 LF	26 SUB	42 *	58 :	74 J	90 Z	106 j	122 z		
11 VT	27 ESC	43 +	59 ;	75 K	91 [107 k	123 {		
12 FF	28 FS	44 ,	60 <	76 L	92 \	108 l	124		
13 CR	29 GS	45 -	61 =	77 M	93]	109 m	125 }		
14 SO	30 RS	46 .	62 >	78 N	94 ^	110 n	126 ~		
15 SI	31 US	47 /	63 ?	79 O	95 _	111 o	127 DEL		

Introduction To Deep Learning

Curse of Dimensionality

Pure Data Space(Pixel Space or Byte Space) is Too big.



$$512 \times 512 \times 3 = 786432$$
$$30**786432 = ?$$

Introduction To Deep Learning

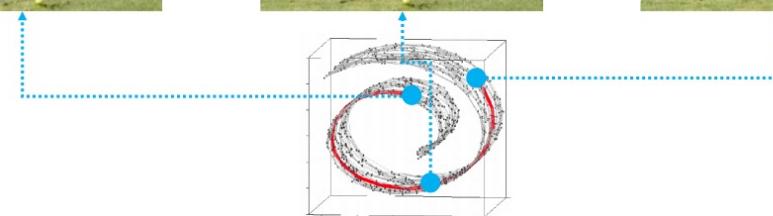
Curse of Dimensionality

Meaningful Data Space is Too small than pure data space.

Reasonable distance metric



Interpolation in manifold



Introduction To Deep Learning

Goal

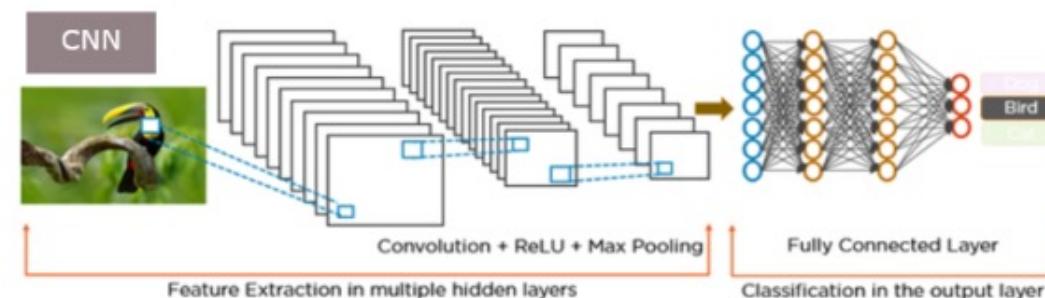
In unstructured data, pure data Space is Too big

- But, meaningful data Space is Too small than Pure data space
- If we can model meaningful data space, we can model unstructured data
- How? We can use some assumption

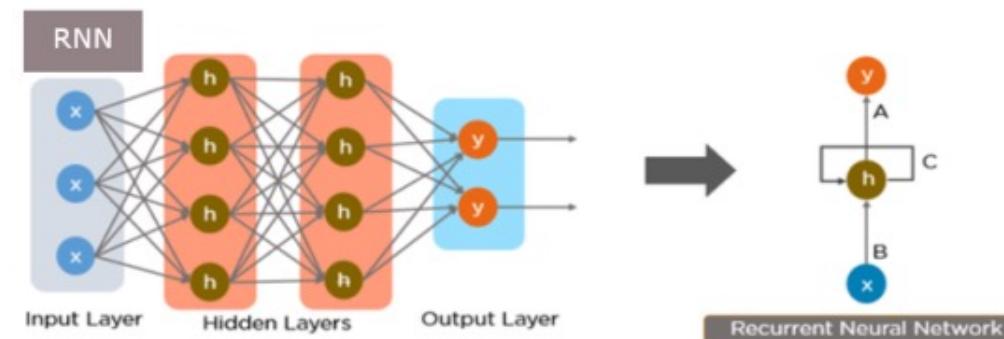
Introduction To Deep Learning

Inductive Bias

Convolutional Neural Network



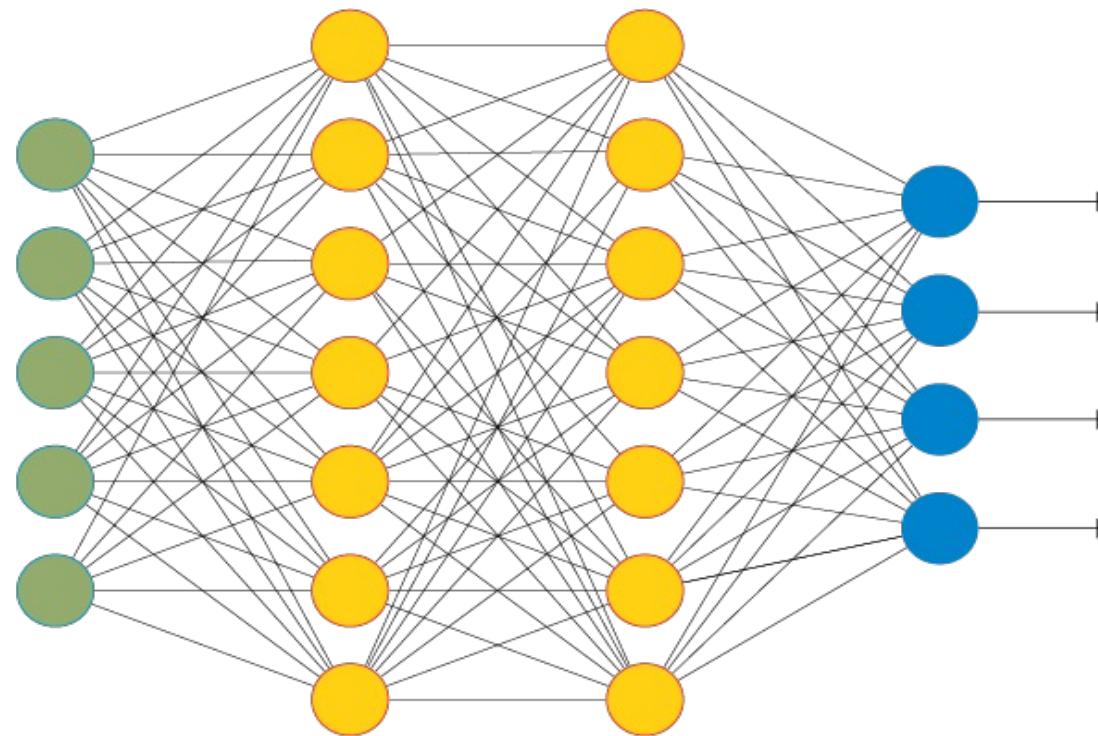
Recurrent Neural Network



Introduction To Deep Learning

Pure Data 2 Meaningful Vector

Vector in Pixel Space
512*512*3 dim Vector



Q11.

정보량의 관점에서 입력 데이터가
정보량이 많을까, 출력 데이터가 정보량이
많을까?

**Vector in Meaningful Space
512 dim Vector
(Manifold Axis)**

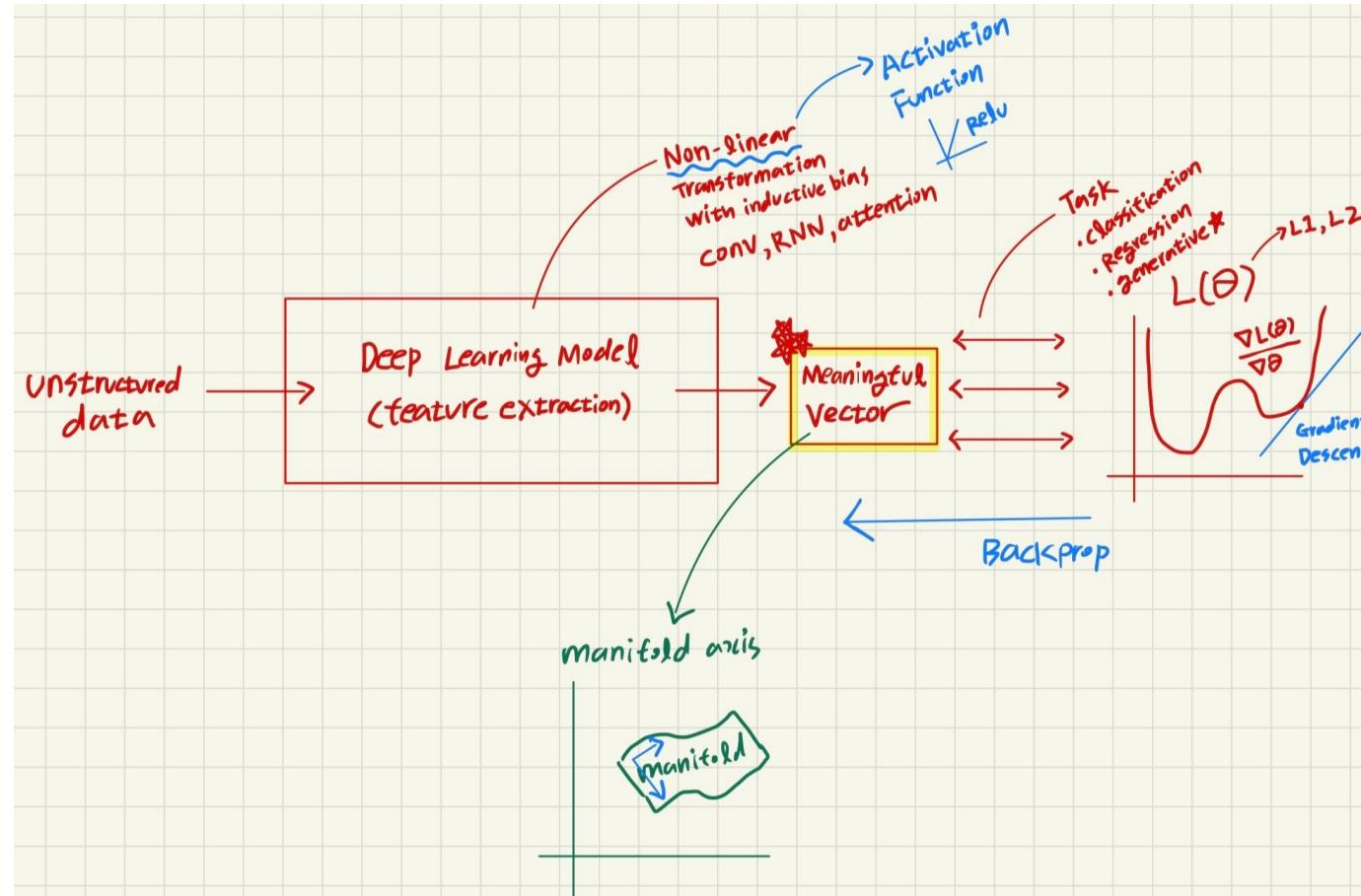
Q12.

데이터가 무한한 상황을 가정하고, 그냥
CNN 없이 학습하면 잘 될까? 분포를
메울만큼 데이터가 없어서 모델링이 잘
안된 것이라면

We can model unstructured data with deep learning

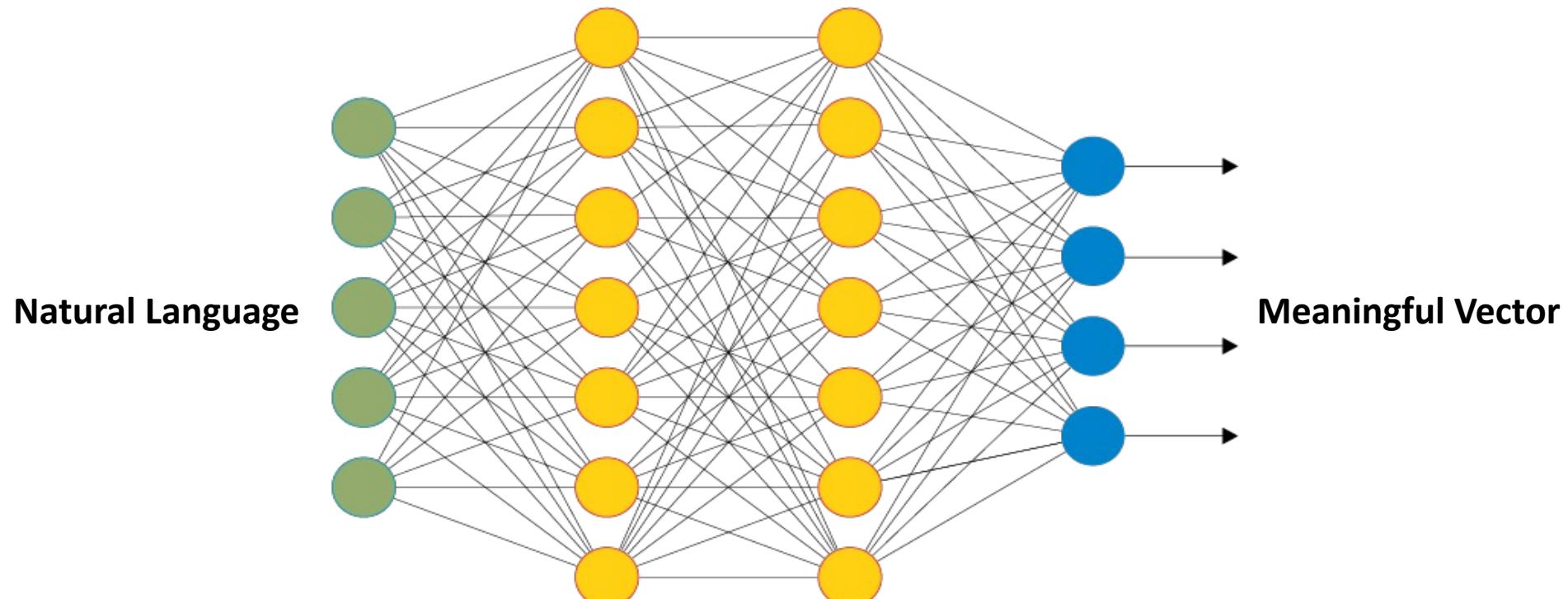
Introduction To Deep Learning

Deep Learning Summary



Introduction to Natural Language Processing

What is NLP

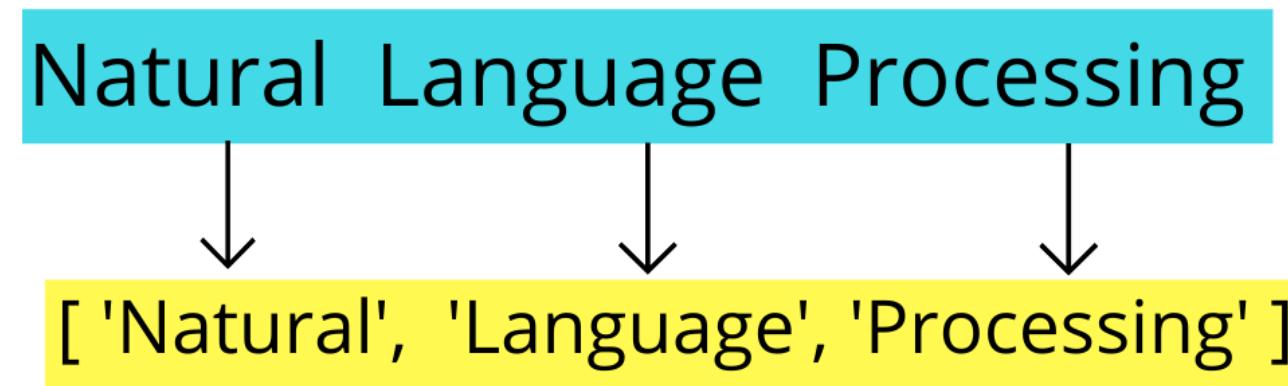


1. How can we get Meaningful Vector
2. Application Using Meaningful Vector

Introduction to Natural Language Processing

How can we get Meaningful Vector

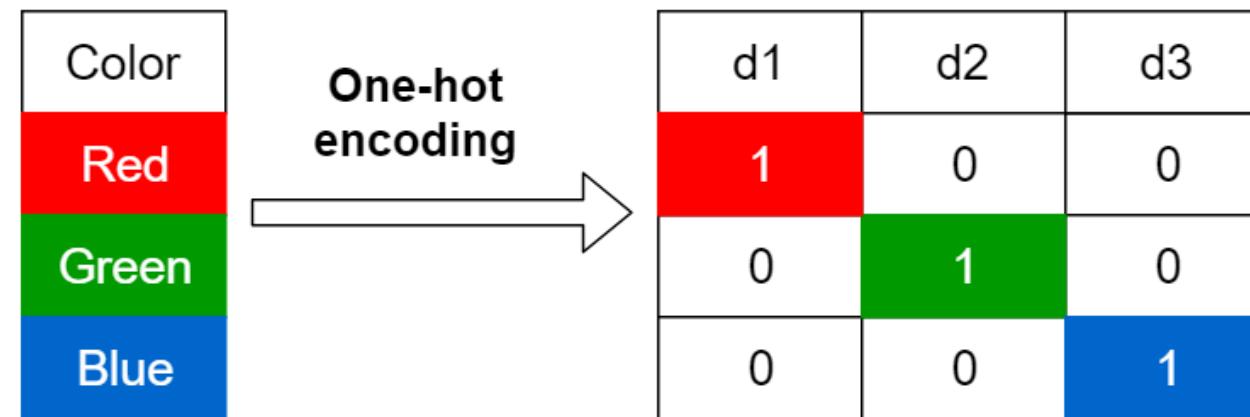
1. Tokenization



Introduction to Natural Language Processing

How can we get Meaningful Vector

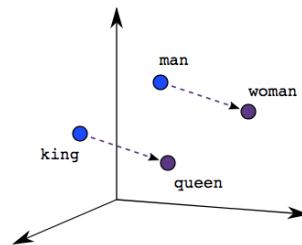
2. One hot Encoding



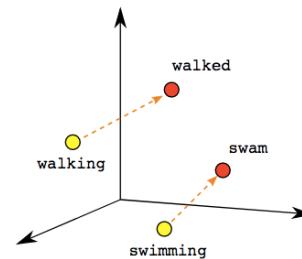
Introduction to Natural Language Processing

How can we get Meaningful Vector

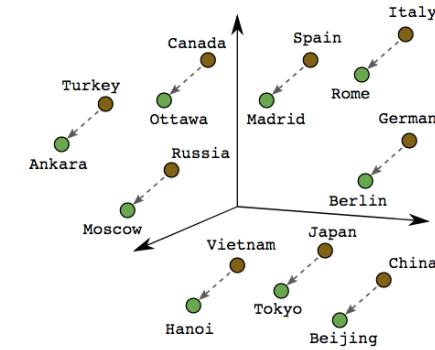
3. Word Embedding



Male-Female



Verb Tense

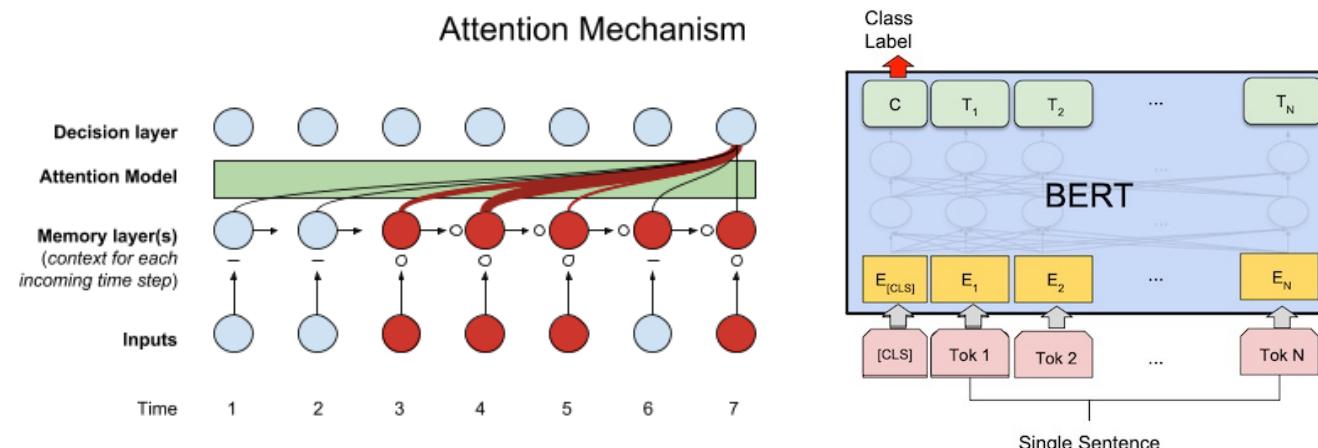
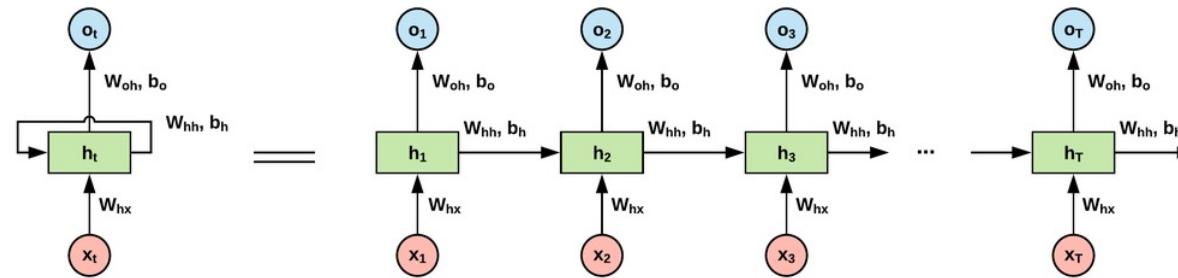


Country-Capital

Introduction to Natural Language Processing

How can we get Meaningful Vector

4. Text Embedding



Introduction to Natural Language Processing

Application Using Meaningful Vector

1. Sentiment Analysis

Sentiment Analysis is the task of classifying the polarity of a given text. For instance, a text-based tweet can be categorized into either "positive", "negative", or "neutral". Given the text and accompanying labels, a model can be trained to predict the correct sentiment.

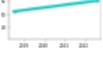
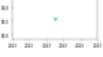
Trend	Dataset	Best Model	Paper	Code	Compare
	SST-2 Binary classification	T5-11B			See all
	IMDb	XLNet			See all
	SST-5 Fine-grained classification	Heinsen Routing + RoBERTa Large			See all
	Yelp Binary classification	XLNet			See all
	MR	VLAWE			See all
	Yelp Fine-grained classification	XLNet			See all
	User and product information	MA-BERT			See all

Introduction to Natural Language Processing

Application Using Meaningful Vector

2. Text Retrieval

Text retrieval in deep learning refers to the process of retrieving relevant textual documents or passages from a large corpus in response to a query or search input. This task is fundamental in various natural language processing (NLP) applications such as information retrieval, question answering, and document summarization.

Trend	Dataset	Best Model	Paper	Code	Compare
	MTEB	SGPT-5.8B-msmarco			See all
	Image-Chat	PaCE			See all
	RSICD	GeoRSCLIP-FT			See all

Introduction to Natural Language Processing

Application Using Meaningful Vector

3. Text Generation

Text Generation is the task of generating text with the goal of appearing indistinguishable to human-written text. This task is more formally known as "natural language generation" in the literature.

T	Model	Average	ARC	HellaSwag	MMLU	TruthfulQA
◆	abacusai/Smaug-72B-v0.1	80.48	76.02	89.27	77.15	76.67
◆	ibivibiv/algaca-dragon-72b-v1	79.3	73.89	88.16	77.4	72.69
💬	moreh/MoMo-72B-lora-1.8.7-DPO	78.55	70.82	85.96	77.13	74.71
◆	cloudyu/TomGrc_FusionNet_34Bx2_MoE_v0.1_DPO_f16	77.91	74.06	86.74	76.65	72.24
◆	cloudyu/TomGrc_FusionNet_34Bx2_MoE_v0.1_full_linear_DPO	77.52	74.06	86.67	76.69	71.32
◆	zhengr/MixTAO-7Bx2-MoE-v8.1	77.5	73.81	89.22	64.92	78.57
💬	yunconglong/Truthful_DPO_TomGrc_FusionNet_7Bx2_MoE_13B	77.44	74.91	89.3	64.67	78.02

Introduction to Natural Language Processing

Application Using Meaningful Vector

4. Text to Img Generation (Unimodal)

Text-to-Image Generation is a task in computer vision and natural language processing where the goal is to generate an image that corresponds to a given textual description. This involves converting the text input into a meaningful representation, such as a feature vector, and then using this representation to generate an image that matches the description.

Trend	Dataset	Best Model	Paper	Code	Compare
	MS COCO	Parti Finetuned			See all
	CUB	TLDM			See all
	Multi-Modal-CelebA-HQ	Swinv2-Imagen			See all
	Oxford 102 Flowers	VQ-Diffusion-F			See all

Introduction to Natural Language Processing

Our Paper Review Curriculum

3/19 <Word Embedding>

Word2Vec : Efficient Estimation of Word Representations in Vector Space (2013)

GloVe : Global Vectors for Word Representation (2014)

ELMo : Deep contextualized word representations (2018)

3/26 <Text Embedding>

RNN : Recurrent neural network based language model (2010)

LSTM : Long Short Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling (2014)

GRU : Learning Phrase Representation using RNN Encoder-Decoder for Statistical Machine Translation (2014)

4/9 <Text Embedding>

Seq2Seq : Sequence to Sequence Learning with Neural Networks (2014)

Attention : Neural Machine Translation by Jointly Learning to Align and Translate (2015)

Transformer : Attention is All You Need (2017)

4/30 <Text Embedding>

BERT Pretraining of Deep Bidirectional Transformers for Language Understanding (2018)

RoBERTa : RoBERTa: A Robustly Optimized BERT Pretraining Approach (2019)

ALBERT: A Lite BERT for Self-supervised Learning of Language Representations (2019)

5/7 <Text Regression/Classification>

MTEB: Massive Text Embedding Benchmark

XLNet: Generalized Autoregressive Pretraining for Language Understanding

5/21 <Text Generation>

GPT-1 : Improving Language Understanding by Generative Pre-Training (2018)

GPT-2 : Language Models are Unsupervised Multitask Learners (2018)

GPT-3 : Language Models are Few-Shot Learners (2020)

5/28 <Document Retrieval>

Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks

Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection

Self-Knowledge Guided Retrieval Augmentation for Large Language Models

6/4 <LLM>

LLaMA: Open and Efficient Foundation Language Models (2023)

Mistral 7B

Tip For Study

Research Follow Up

1. Hugging Face Daily Papers

The screenshot shows the Hugging Face Daily Papers interface. At the top, there is a navigation bar with links for Models, Datasets, Spaces, Posts, Docs, Pricing, Log In, and Sign Up. A search bar is also present. Below the navigation bar, there is a banner with the text "Get trending papers in your email inbox once a day!" and a "Subscribe" button. The main content area is titled "Daily Papers" and features a search bar for "arxiv id or title". The date "MAR 8" is displayed in a calendar icon.

Yi: Open Foundation Models by 01.AI

by AK

Abstract

We introduce the Yi model family, a series of language and multimodal models that demonstrate strong multi-dimensional capabilities. The Yi model family is based on 6B and 34B pretrained lanuage models, then we extend them to chat models.

Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference

Wei-Lin Chiang¹, Liamin Zheng², Ying Sheng², Anastasios N. Angelopoulos¹, Tianle Li¹, Dacheng Li¹, Banghua Zhu¹, Hao Zhang², Michael I. Jordan¹, Joseph E. Gonzalez¹, Ion Stoica¹

Abstract

Large Language Models (LLMs) have unlocked new capabilities and applications; however, evaluating the alignment with human preferences still poses significant challenges. To address this issue, we introduce Chatbot Arena, an open platform for evaluating LLMs based on human preference. Our methodology employs a pairwise comparison approach and leverages input from a diverse user base through crowdsourcing. The platform has been operational for several months, involving over 7,000 users. This paper describes

Figure 1. Classification of LLM benchmarks: We categorize along two dimensions: whether the questions are from a static dataset or a live fresh source, and whether the evaluation metric reflects an agreement of the human and machine preferences. MMLU (Hruskova et al., 2020), HellSWAG (Zellers et al., 2019), GSM-8K (Cobbe et al., 2021), MT-Bench (Zheng et al., 2023b), and AlpacaEval (Li et al., 2023) are common examples of static benchmarks. Chatbot Arena is the platform introduced in this paper.

Teaching Large Language Models to Reason

Tip For Study

Research Follow Up

2. Papers with Code

Search  Browse State-of-the-Art Datasets Methods More 

  Sign In

 Natural Language Processing

Sentiment Analysis

1270 papers with code • 43 benchmarks • 91 datasets

Sentiment Analysis is the task of classifying the polarity of a given text. For instance, a text-based tweet can be categorized into either "positive", "negative", or "neutral". Given the text and accompanying labels, a model can be trained to predict the correct sentiment.

Sentiment Analysis techniques can be categorized into machine learning approaches, lexicon-based approaches, and even hybrid methods. Some subcategories of research in sentiment analysis include: multimodal sentiment analysis, aspect-based sentiment analysis, fine-grained opinion analysis, language specific sentiment analysis.

More recently, deep learning techniques, such as RoBERTa and T5, are used to train high-performing sentiment classifiers that are evaluated using metrics like F1, recall, and precision. To evaluate sentiment analysis systems, benchmark datasets like SST, GLUE, and IMDB movie reviews are used.

Further readings:

- [Sentiment Analysis Based on Deep Learning: A Comparative Study](#)

Benchmarks

Add a Result

These leaderboards are used to track progress in Sentiment Analysis

Trend Dataset Best Model Paper Code Compare



Content

- Introduction
- Benchmarks
- Datasets
- Subtasks
- Libraries
- Papers
 - Most implemented
 - Social
 - Latest
 - No code

Tip For Study

Research Follow Up

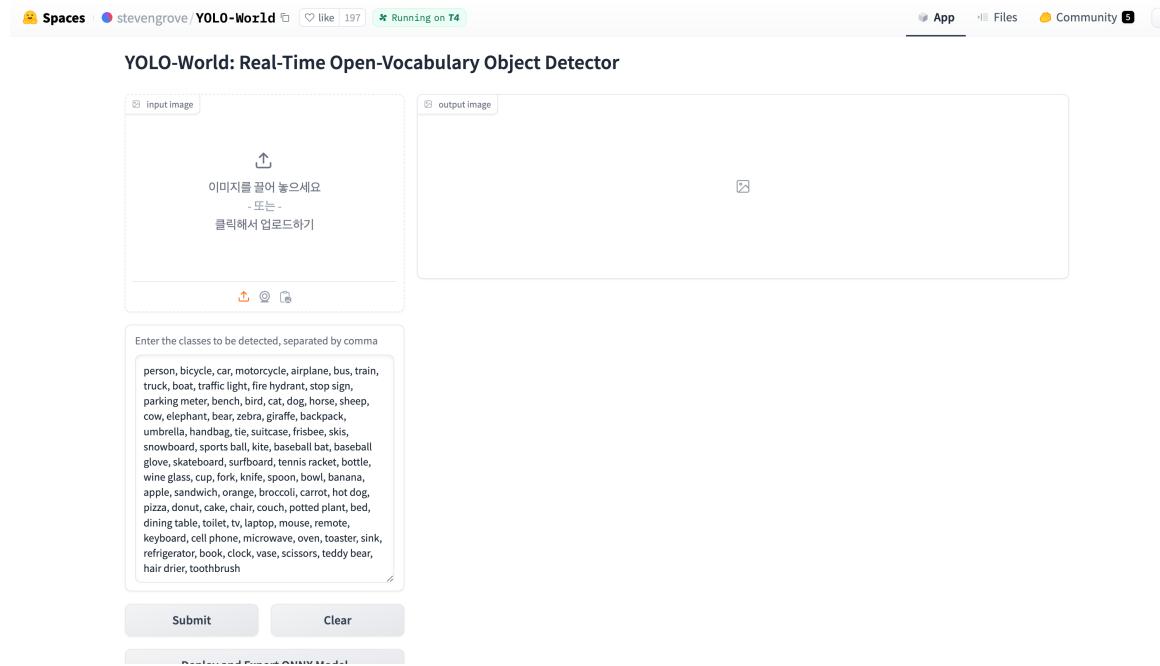
3. Github Awesome 000

The screenshot shows a GitHub repository page for 'keon / awesome-nlp'. The repository is public and has 609 watchers, 2.6k forks, and 15.8k stars. It contains 4 branches and 0 tags. The main page displays a list of commits from NirantK, including merging pull requests and updating files like README, LICENSE, and PULL_REQUEST_TEMPLATE. The repository is described as a curated list of resources dedicated to NLP, with tags including nlp, language, machine-learning, natural-language-processing, text-mining, awesome, deep-learning, and awesome-list. The README file indicates CC0-1.0 license. The repository has no releases or packages published.

Tip For Study

Code Demo

4. Hugging Face Gradio Demo



Tip For Study

Paper Review

5. Notion

Paper Review

arXiv 20xx. [Paper]
Authors
Affiliation
1 Jan 2024

Summary
Task :
Method
Pipeline
Discussion

Summary

Task :

Method

Pipeline

Discussion



T R A I N A N D T E S T