

Generative Model & Multimodal Model

Hyunbeen Kim

been000904@gmail.com

NLP Team

2024/09/12

Contents

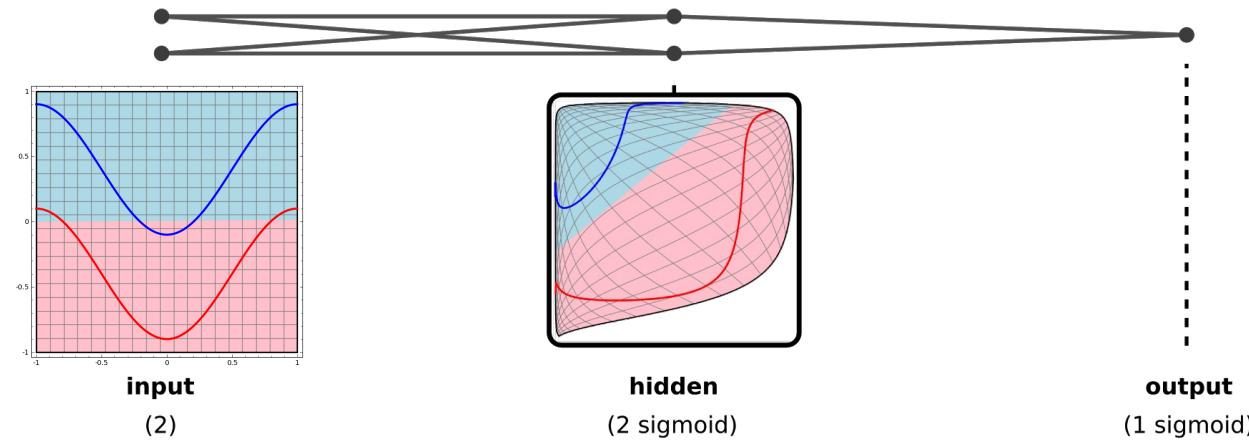
- NLP Team Rule
- DL Review
- Generative Model
- Multimodal Model
- Tip for Study
- Our plan

NLP Team Rule

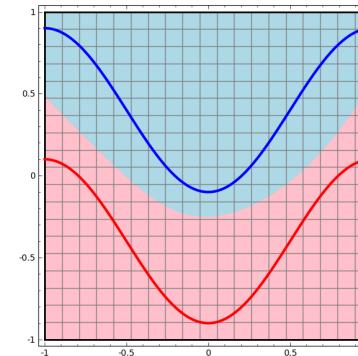
1. 매주 화요일 19시~21시 사당역 Moim에서 진행
2. 커리큘럼은 주제와 관련된 논문 리뷰하는 방식으로 진행
3. 한두명의 발제자가 한 주제와 관련한 논문 3편 정도 리뷰
4. 발제자가 아니더라도 웬만하면 다음 진행할 논문은 다 읽어오기 !
5. 디스커션시 과대망상 환영 !
6. 프로젝트는 자유롭게 진행 ! 함께 하고 싶으면 팀에 제안하는 식으로

Deep Learning Review

How to Solve Non-Linear Problem (Solve non-Linear Separable problem)

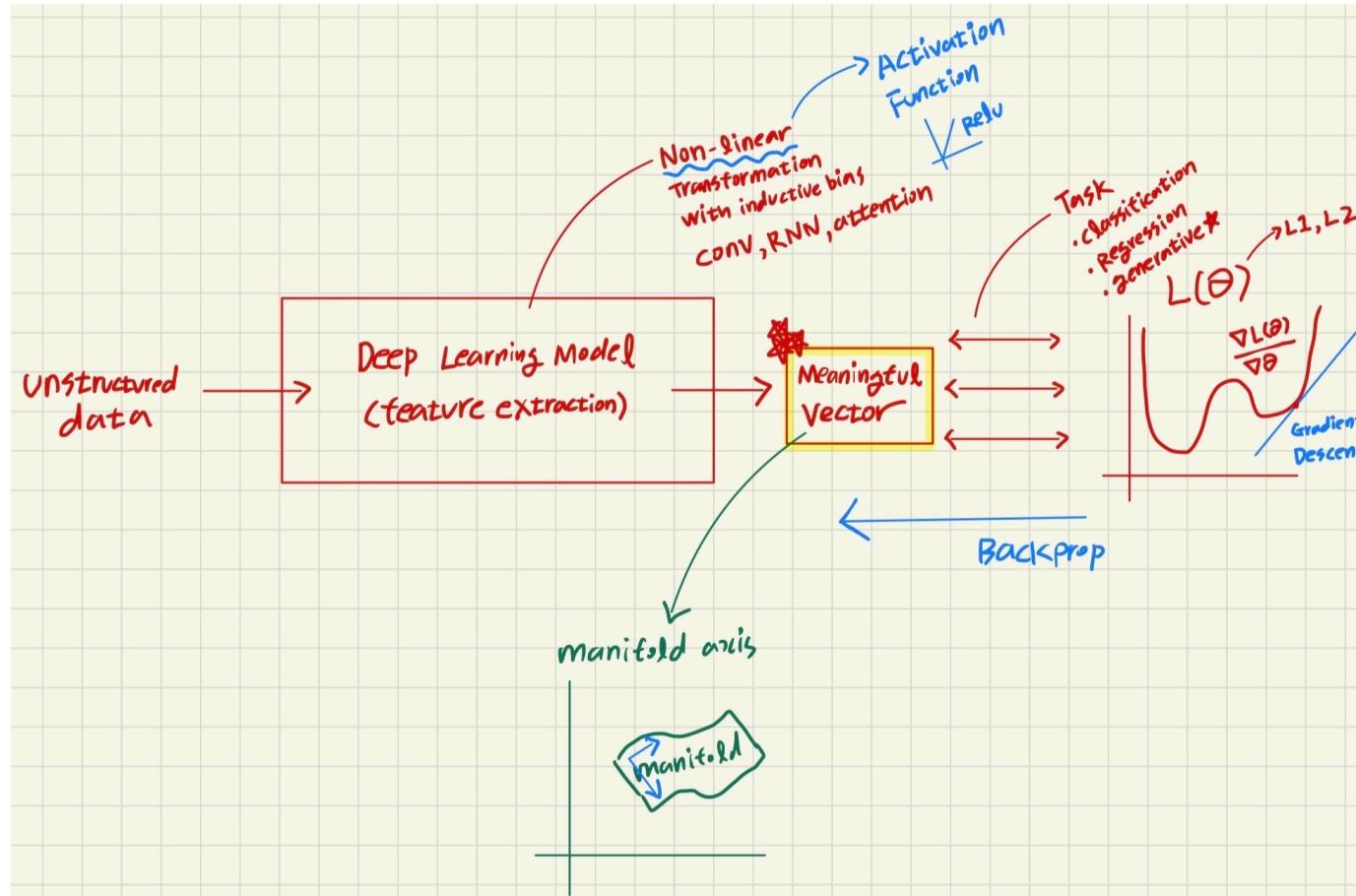


Result:



Deep Learning Review

Deep Learning Summary



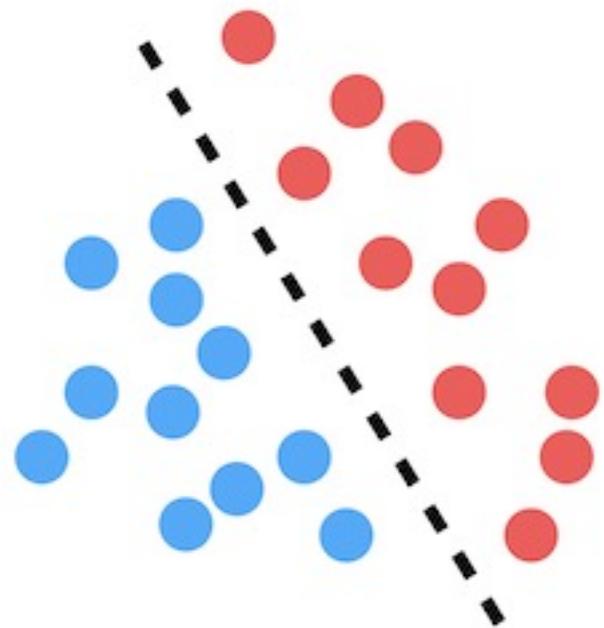
Generative Model

What is Generative Model?

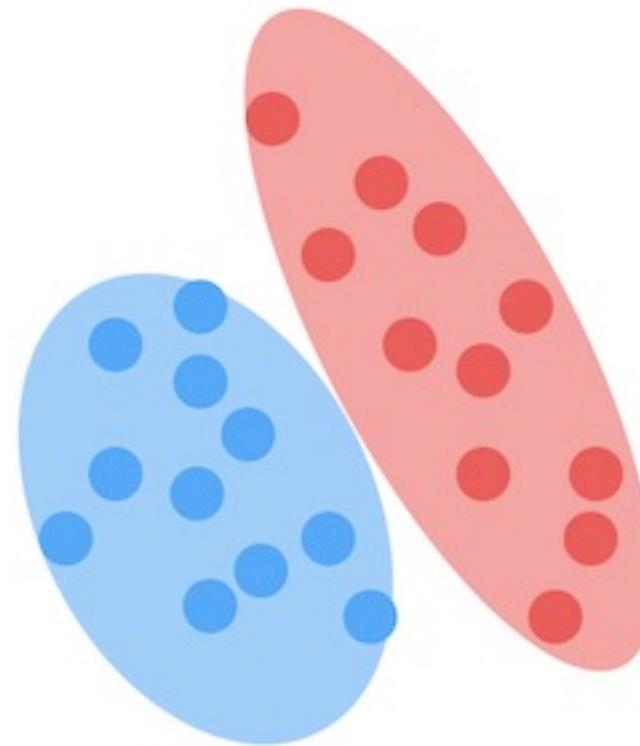
Generative Model

Discriminative Model VS Generative Model

Discriminative



Generative



Generative Model

Discriminative model approximates $P(Y|X)$

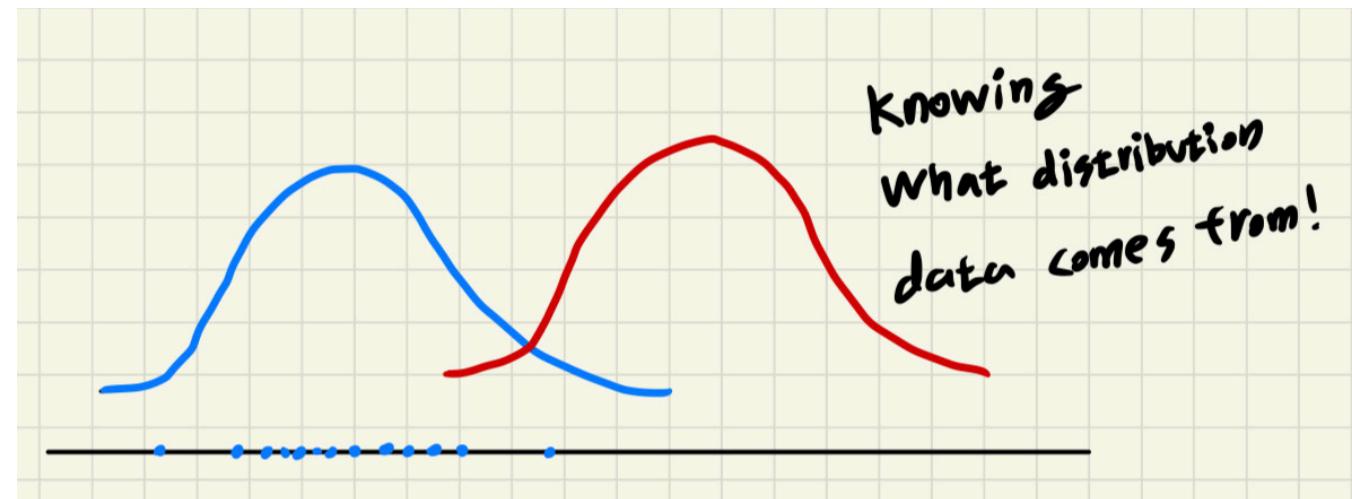
/

Generative model approximates $P(X)$

오늘 세미나에서 가장 중요한 내용!

Generative Model

$P(x)$ 를 알고 있다는 것은 무엇을 의미할까?



Generative Model

$P(X)$ 를 알면 뭐할 수 있는데?

데이터 분포 $P(X)$ 에서 새로운 데이터를 뽑아 사용할 수 있음

=> Stable Diffusion, ChatGPT and DallE

우도를 기반으로 데이터의 클래스나 이상치를 식별할 수 있음

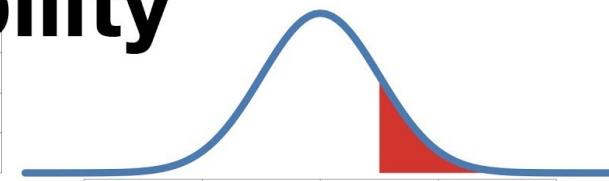
=> Discriminative Model처럼 분류 태스크가 가능해짐

Generative Model

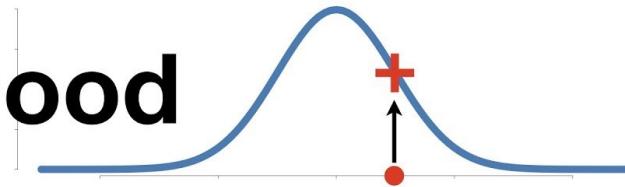
잠만 근데 우도가 뭐지?

Probability

Vs



Likelihood



Probability Distribution Function의 output 값

Generative Model

$P(x)$ 를 아는 것이 좋은 이유는 알겠어?

근데 그걸 어떻게 구해?

케이스에 따라 다른데 $P(x)$ 를 구하는게(근사하는게)

굉장히 어려울 때가 있고, 굉장히 쉬울 때가 있음

Generative Model

먼저 $P(x)$ 를 구한다는 걸 조금 더 엄밀하게 정의할 수 있어야 하지 않나?

암시적 생성 모델(Implicit Model)에선 $P(x)$ 에서
뽑은 데이터를 얻을 수 있으면 $P(x)$ 를 구하는 것이라고 정의
(Ex. Generative Adversarial Network)

VS

명시적 생성 모델(Explicit Model)에선 $P(x)$ 를
Likelihood를 구할 수 있으면 $P(x)$ 를 구한 것이라고 정의
(Ex. Autoregressive Model, Diffusion Model)

Generative Model

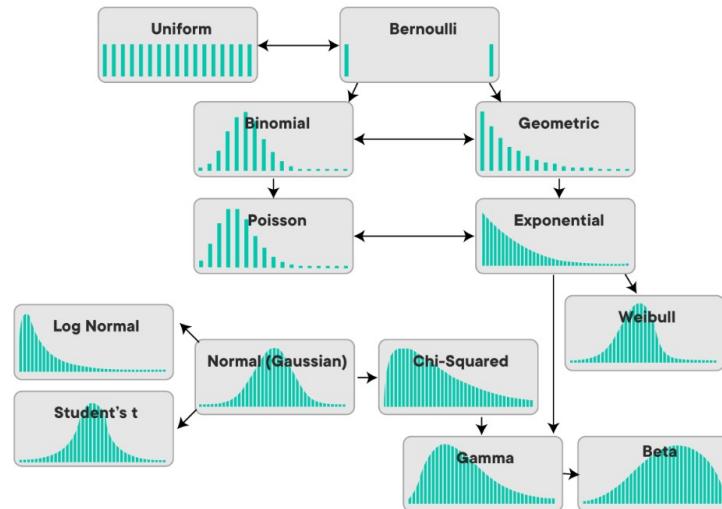
잠시 흐름 정리

생성 모델은 $P(x)$ 를 근사하는 모델

- ⇒ $P(x)$ 를 안다는 것은 데이터가 온 것으로 추정되는 분포의 종류와 분포의 모수를 아는 것
 - ⇒ 이걸 알면 샘플링을 할 수도 있고, 우도 기반으로 데이터 분류도 가능해짐
 - ⇒ 우도를 구할 수 있거나, 데이터를 샘플링 할 수 있을 때 $P(x)$ 를 안다고 이야기할 수 있음

Generative Model

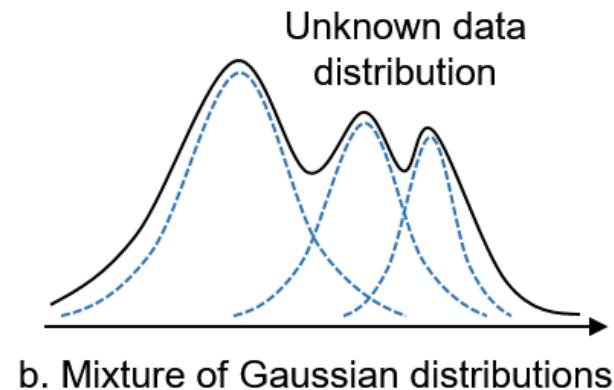
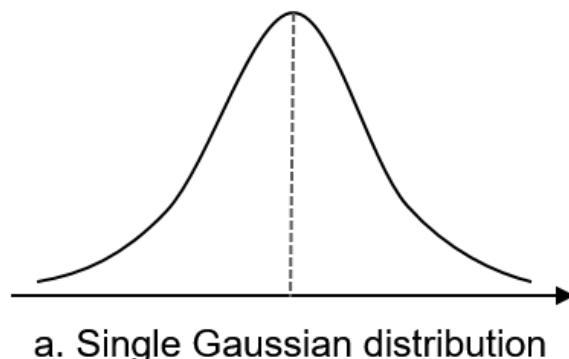
데이터가 간단한 분포에서 왔다고 가정할 수 있을 때 $P(x)$ 를 안다는 것의 의미



데이터가 나온 분포의 종류 (Gaussian, Laplacian and Bernoulli)를 알고 있고, 분포의 모수를 아는 것,
이런 경우 사실 샘플링도 우도를 구하는 것도 모두 쉽게 가능

Generative Model

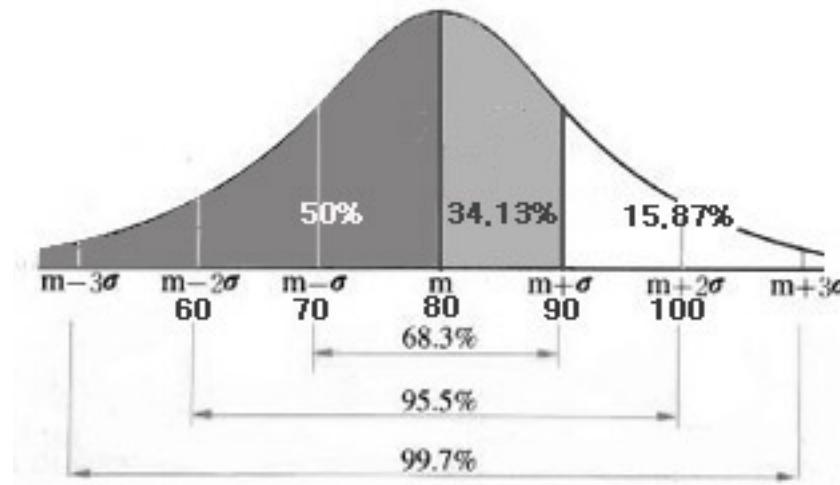
데이터가 복잡한 분포에서 왔을 것으로 보일 때 $P(x)$ 를 안다는 것의 의미



분포의 종류를 알 수 없으니 분포에서 온 데이터를 어떤 방식으로든 만들 수 있을 때나
여러 비선형 변환으로 우도를 구할 수 있을 때 $P(x)$ 를 안다고 말할 수 있음

Generative Model

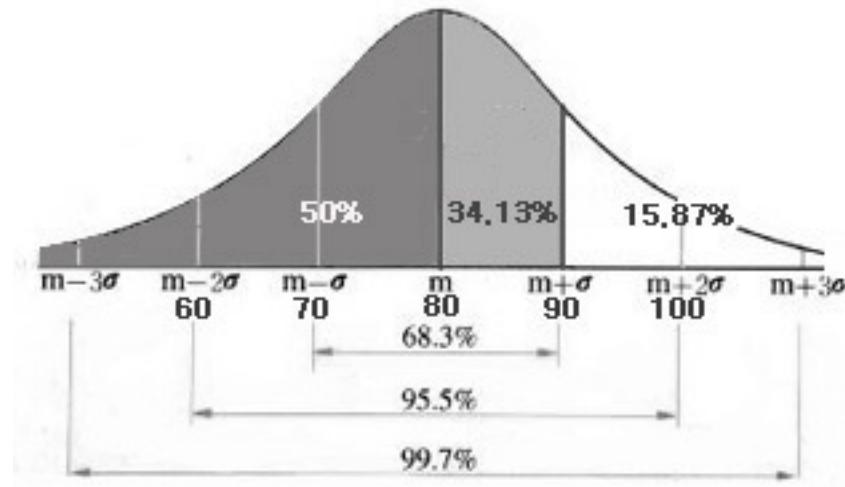
$P(x)$ 를 구하는게 쉬운 케이스 예시: 학생 성적



일반적으로 성적은 정규분포를 따른다고 가정할 수 있음, 정규분포의 모수는 평균과 표준편차임으로 주어진 데이터가 있을 때 충분히 구할 수 있음

Generative Model

$P(x)$ 를 구하는게 쉬운 케이스 예시

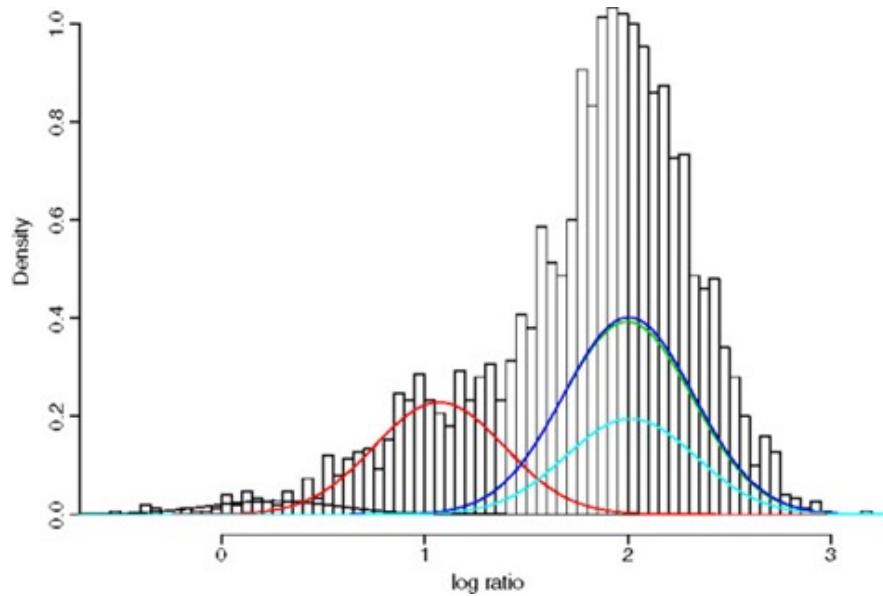


만들어낸 분포를 통해 그럴싸한 학생 성적도 얼마든지 만들 수 있고 (샘플링 가능)

정규분포 식을 알기 때문에 PDF의 output인 우도도 당연히 구할 수 있음

Generative Model

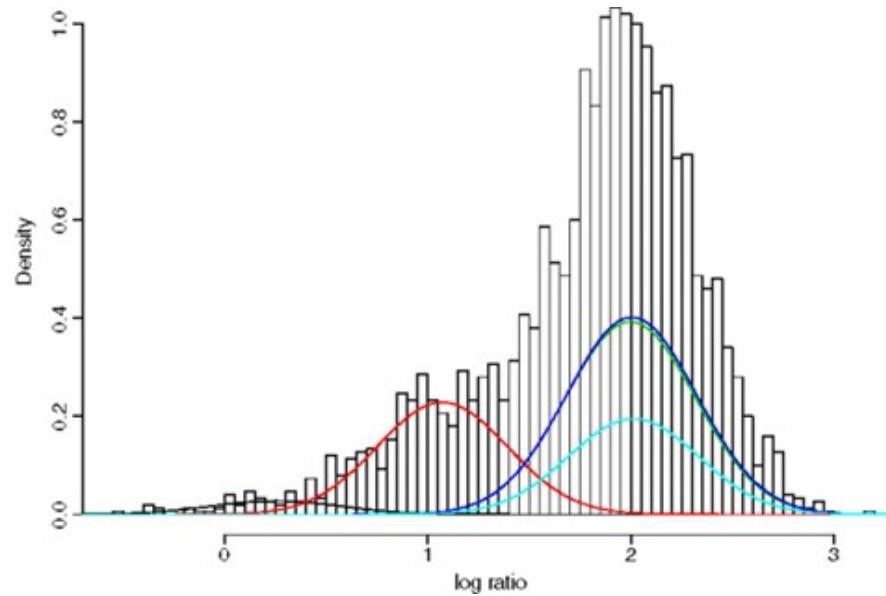
$P(x)$ 를 구하는게 살짝 어려운 케이스 예시: 남여 키가 섞인 분포



간단한 분포로 설명은 못할 것 같은데, 간단한 분포를 몇개 섞으면
잘 설명할 수 있을 것 같은 데이터

Generative Model

$P(x)$ 를 구하는게 살짝 어려운 케이스 예시: 남여 키가 섞인 분포

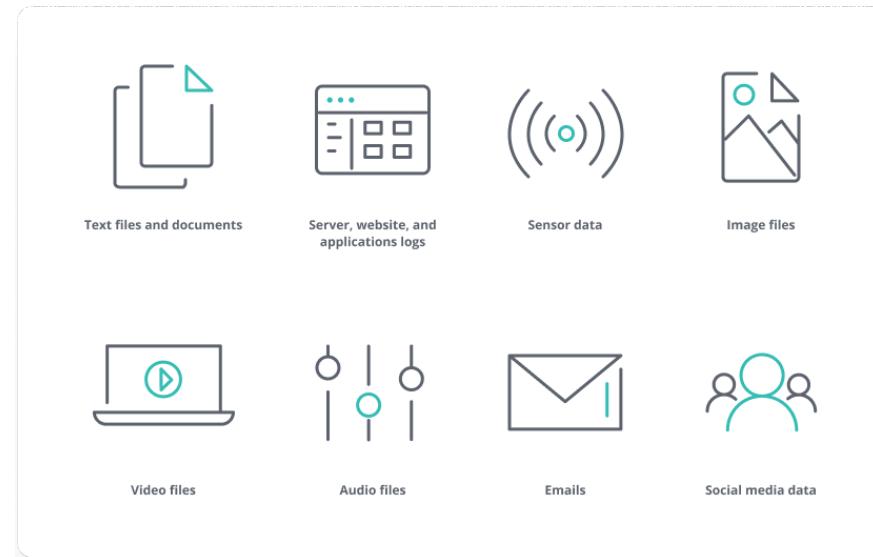


샘플링 할 때마다 어떤 분포에서 뽑을 것인지 랜덤하게 결정하면 되니까

샘플링 당연히 가능하고, 우도도 당연히 정할 수 있음

Generative Model

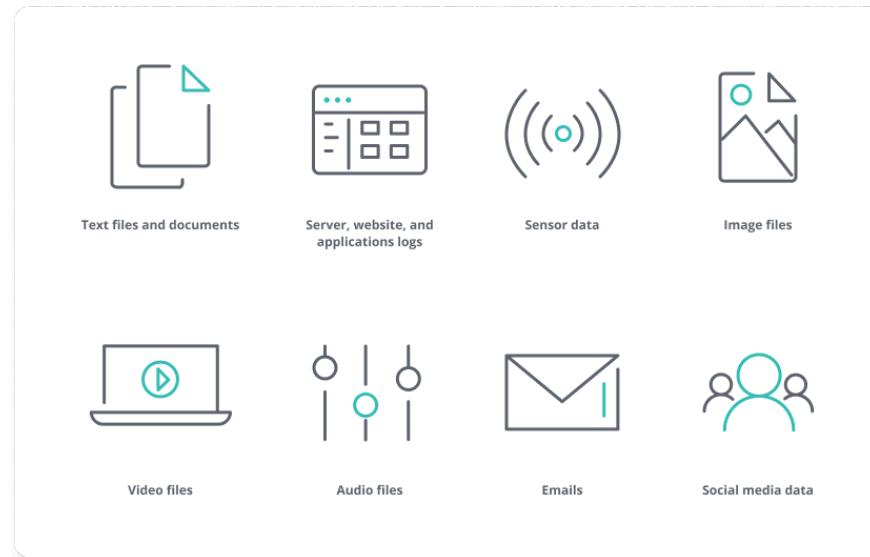
$P(x)$ 를 구하는게 굉장히 어려운 케이스 예시: 비정형 데이터



직관적으로 분포를 떠올릴 때 x 축에 무엇을 두어야 할지도
감이 안오는 굉장히 고차원의 데이터

Generative Model

$P(x)$ 를 구하는게 굉장히 어려운 케이스 예시: 비정형 데이터

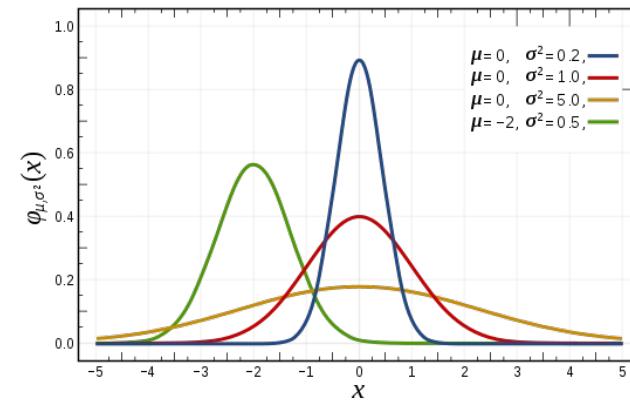
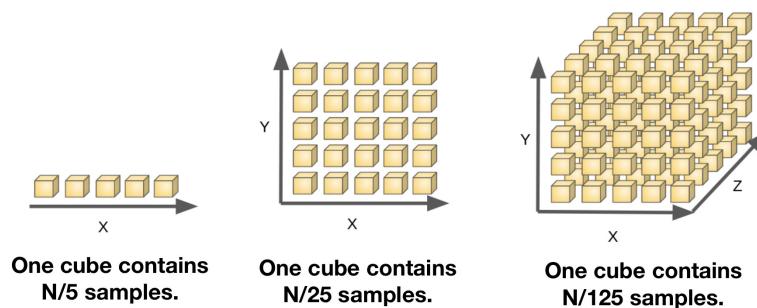


일반적인 방법으로는 우도를 구하기도 샘플링하기도 쉽지 않음

이번 학기 우리가 풀어나갈 문제는 앞선 두 문제가 아닌 지금 이 문제

Generative Model

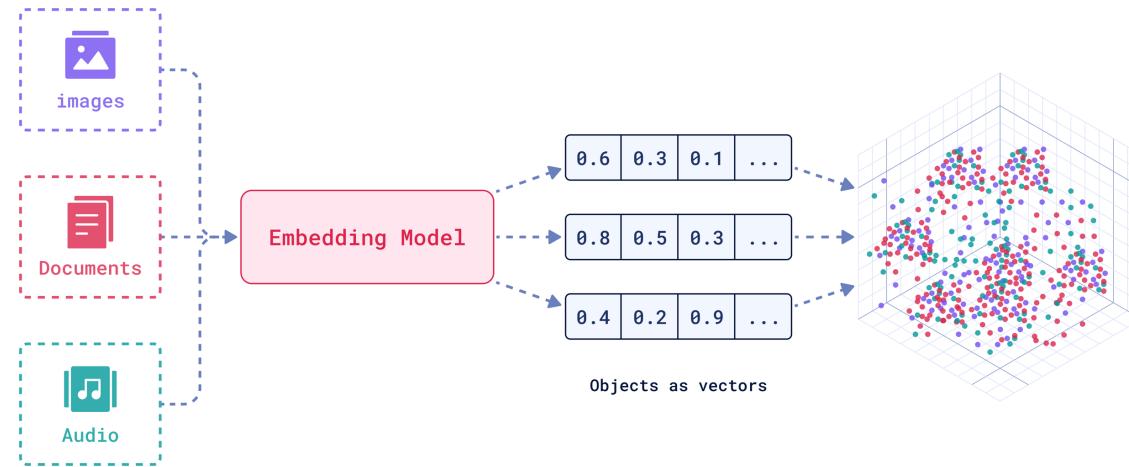
비정형 데이터에 대한 $P(x)$ 구하기는 왜 어려울까?



1. 모델링하기에 차원이 너무 크다.
2. 근사하고자 하는 분포가 너무 복잡하다.

Generative Model

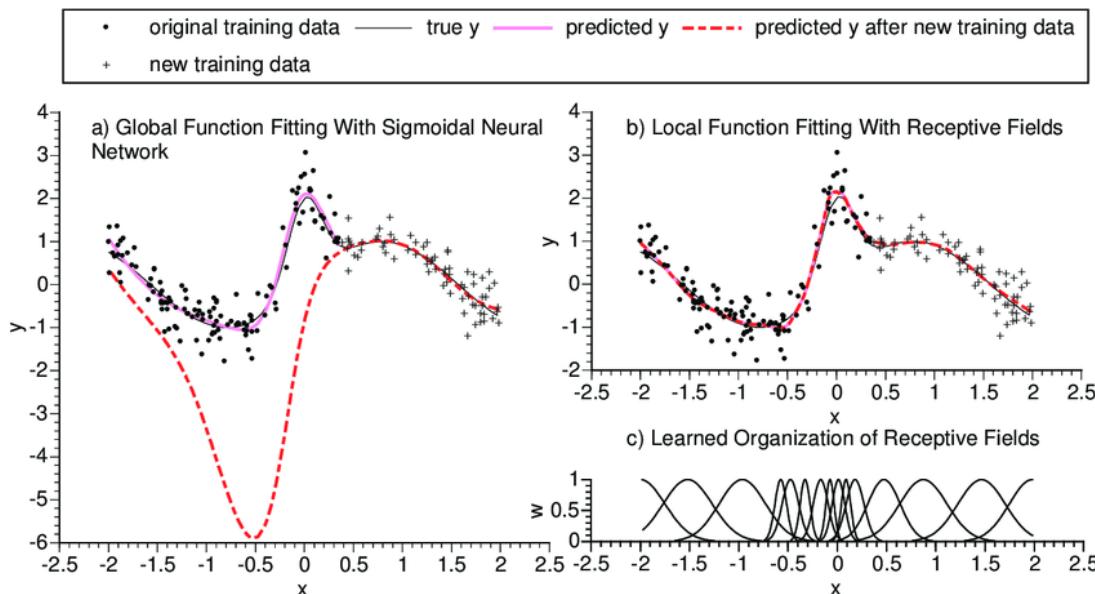
차원을 어떻게 하면 줄일 수 있을까?



임베딩을 통한 정보 압축 => Deep Learning이 제일 잘하는거

Generative Model

근사하고자 하는 분포가 너무 복잡다는 문제는?



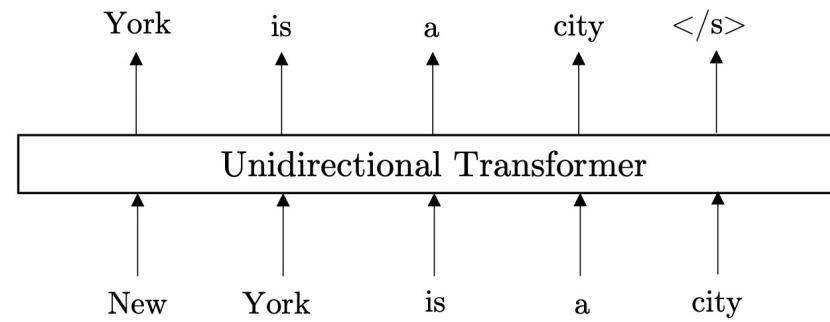
함수를 근사하는 것 => Deep Learning이 제일 잘하는거2

Generative Model

결국 근사할 함수만 잘 정의되면 비정형 데이터 생성 모델도 충분히 가능
수학적으로 Deep Learning은 유한 파라미터를 가진 모든 함수를 근사할 수 있음

Generative Model

첫번째 방법: Autoregressive Model



$$\log p(\mathbf{x}) = \sum_{t=1}^T \log p(x_t | \mathbf{x}_{<t})$$

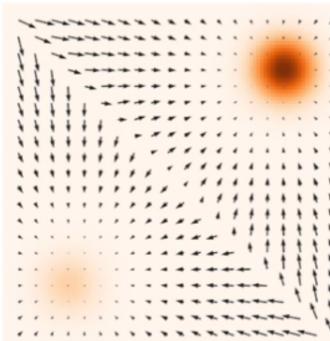
T-1까지의 output을 기반으로 T번째에 나올 값 예측하는 함수 근사
첫번째 데이터를 랜덤으로 뽑으면 다음 데이터부터는 모델이 예측

Generative Model

두번째 방법: Diffusion Model

$$\frac{1}{2} \mathbb{E}_{q_\sigma(\tilde{\mathbf{x}}|\mathbf{x}) p_{\text{data}}(\mathbf{x})} [\|\mathbf{s}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}} | \mathbf{x})\|_2^2].$$

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{x}}_{t-1} + \frac{\epsilon}{2} \nabla_{\mathbf{x}} \log p(\tilde{\mathbf{x}}_{t-1}) + \sqrt{\epsilon} \mathbf{z}_t,$$



(pdf and score)

PDF의 기울기를 구하는 함수 근사

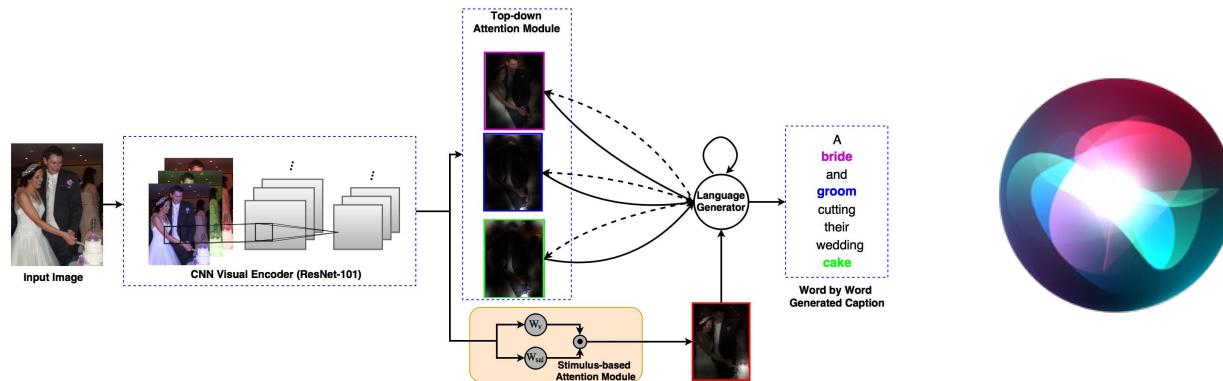
PDF의 기울기를 따라 올라가며 높은 우도를 가진 데이터 샘플링

Multimodal Model

What is Multimodal Model?

Multimodal Model

Image2Text, Text2Speech, Speech2Text?



Nah...

Multimodal Model

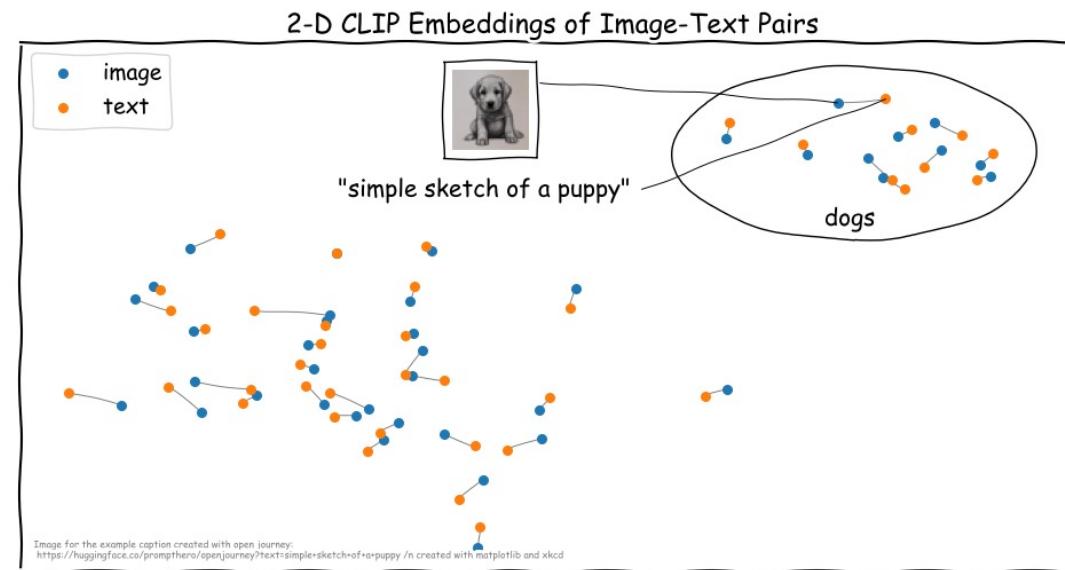
멀티모달 모델이면 여러 모달의 인풋에 대해 이해할 수 있어야 함

=>

이는 여러 모달의 인풋을 하나의 임베딩 스페이스로 보낼 수 있어야 함 의미

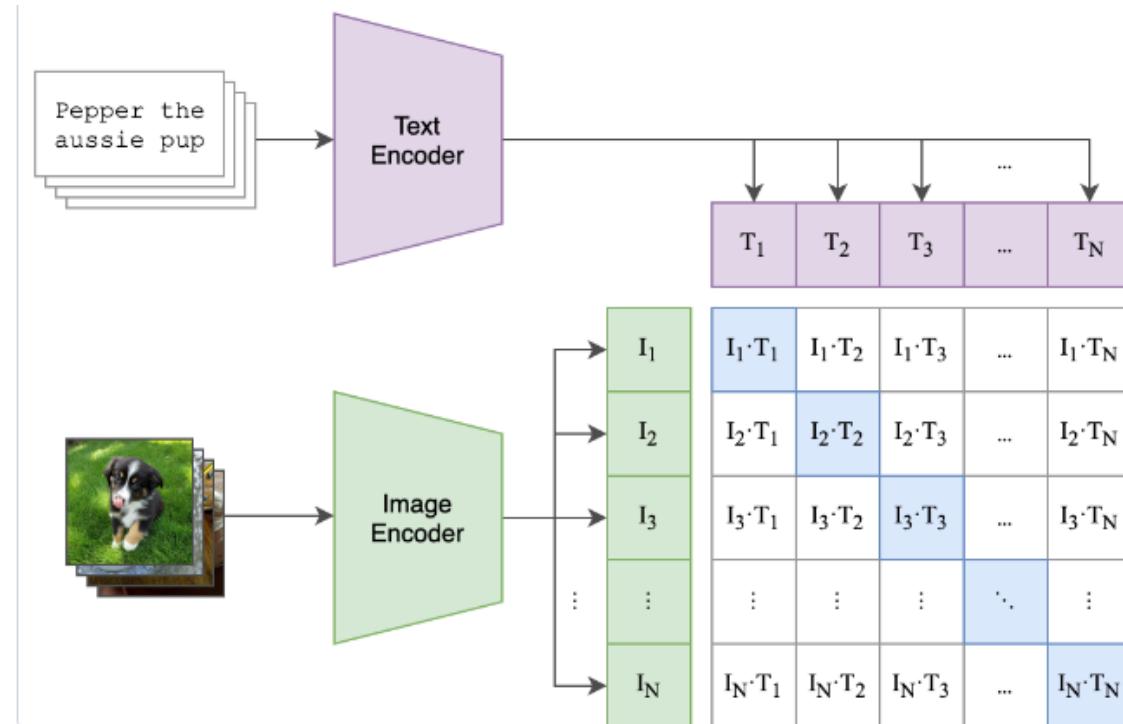
=>

여러 모달에 대한 임베딩이 연산이 가능해야 함



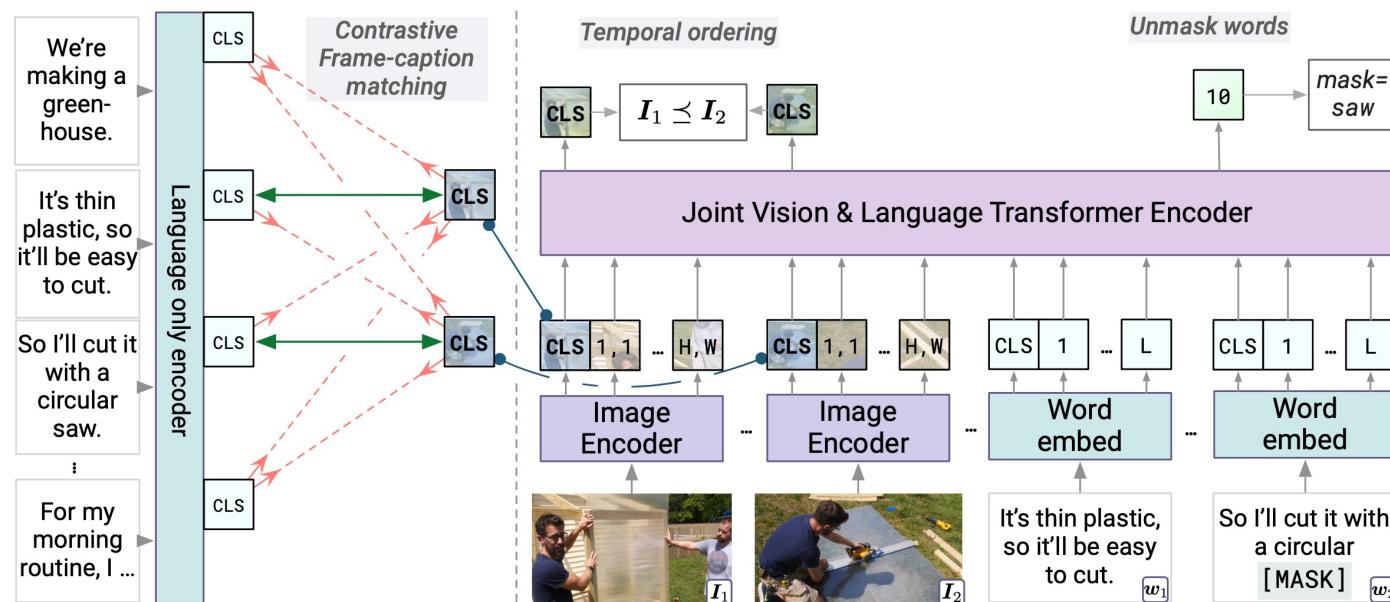
Multimodal Model

멀티모달의 가장 기본이 되는 CLIP은 이미지와 텍스트 사이의 거리를 쟈 수 있음



Multimodal Model

최근 많이 연구되었던 VLM도 Vision Info를
조건으로 줄 수 있으니 멀티모달 모델의 범주에 들어갈 수 있음



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". A calendar icon indicates it is March 8th.

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 laumerae 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 2,000 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 on general intent or specific domain-specific 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 mentions CC0-1.0 license. The repository has no releases or packages published.

Our Plan

Our Paper Review Curriculum

Generative AI & Multimodal AI Paper

9/10

Introduction to Generative Model & Multimodal Model

9/24 <Multimodal Representation & Visual Captioning>

CLIP

BLIP

10/1 <Diffusion>

DDPM

NCSN

Stable Diffusion

10/29 <VLM>

LLaVA

LLaVA 1.5

LLaVA 1.6

Idefics2

11/5 <LLM Agent>

Toolformer

Visual ChatGPT

HuggingGPT

Our Plan

Project Open Discussion

공모전?

토이 프로젝트?

주제?

얻어가고 싶은 것?



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