

- Statistic/Math

1 조건부 확률은 무엇일까요?

2 공분산과 상관계수는 무엇일까요? 수식과 함께 표현해주세요.

- Machine Learning

3 LSA, LDA, SVD 등의 약자들이 어떤 뜻이고 서로 어떤 관계를 가지는지 설명할 수 있나요?

4 Markov Chain을 고등학생에게 설명하려면 어떤 방식이 제일 좋을까요?

- Deep Learning

5 볼츠만 머신은 무엇인가요?

6 Tensorflow, PyTorch 특징과 차이가 뭘까요?

7 TF, PyTorch 등을 사용할 때 디버깅 노하우는?

8 뉴럴넷의 가장 큰 단점은 무엇인가? 이를 위해 나온 One-Shot Learning은 무엇인가?

- Python

9 What is type conversion in Python?

- Algorithm

10 Problem

1. 조건부 확률: 어떤 사건 A가 발생했을 때, 사건 B가 발생할 확률

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

2. 공분산: 2개의 확률 변수의 선형 관계를 나타내는 값

$$\text{Cov}(X, Y) = E\{(X - E(X)) \cdot (Y - E(Y))\} = E(X \cdot Y) - E(X) \cdot E(Y)$$

If) X와 Y가 독립이면, $\text{Cov}(X, Y) = 0$

상관 계수: 두 변수 사이의 통계적 관계를 표현하기 위해 특정한 상관 관계의 정도를 수치적으로 나타낸 계수

$$\text{피어슨 상관 계수}(r) = \frac{\text{Cov}(X, Y)}{s_x \cdot s_y} = \frac{s_{xy}}{s_x \cdot s_y}$$

• 상관 계수는 항상 -1과 1 사이에 있다.

• 상관 계수의 절대값 크기는 직선관계에 가까운 정도를 나타내고, 부호는 직선관계의 방향을 나타낸다.

• 상관 계수의 단위는 없다. \Rightarrow 매번 다른 scale의 data를 비교하게 용이하다.

3. SVD(특이값 분해)

• PCA는 데이터의 공분산 행렬을 기반으로 고유벡터를 생성하고 이를 바탕으로 고유벡터에 입력 데이터를 선형 변환하여 차원을 축소하는 방법

• SVD는 PCA와 유사한 행렬 분해 기법을 사용하나 정방행렬을 분해하는 PCA와 달리 행과 열의 크기가 다른 행렬에도 적용할 수 있다.

LSA (잠재 의미 분석)

• 토픽 모델링에서 자주 사용되는 기법으로 DTM이나 TF-IDF 행렬에 Truncated SVD(절단된 SVD)를 적용하여 차원을 축소시키고, 단어들의 잠재적인 의미를 이끌어낸다.

• Truncated SVD는 SVD와 똑같은 상위 n개의 특이값만 사용하는 축소 방법이다. \Rightarrow 복원 불가능

LDA (잠재 디리클레 할당)

• LSA와 달리 단어 특정한 토픽에 존재할 확률과 문서에 특정 토픽이 존재할 확률을 결합 확률로 토픽을 추정하는 방법을 의미

4. Markov Chain: Markov 성질을 가진 이산 확률과정

↳ 특정 상태의 확률은 오직 과거 상태에 의존한다.

즉, 충분히 반복된 특정 시점의 확률로 다음 확률을 예측 (수학적 귀납법)

5. 볼츠만 머신 (Boltzmann machine): 가시 층과 은닉층의 연결만이 남아 있는 형태

모든 뉴런이 연결되어 있는 완전 연결 그래프 형태이며, 제한된 볼츠만 머신에서는 같은 층의 뉴런들 연결되어 있지 않은 모양이다.

기본적으로 단층 구조이며, 확률 모델이다. 분류나 선형 회귀 분석 등에 사용될 수 있다.

특히, DBN (Deep Belief Network)에서는 RBM들을 쌓아올려, 각 볼츠만 머신을 순차적으로 학습시킨다.

6.

구분	Tensorflow	Pytorch
구현 방법	Define and Run	Define by Run
성질	범용적	직관적, 유기적
Graph 형태	Static Graph	Dynamic Graph

7. 디버깅 노하우

- 오류가 발생하는 곳, 중요한 데이터가 바뀌는 지점을 디버깅 포인트로 둔다.
- logging tool (tensorboard, wandb)를 이용한다.
- 오류에 대한 대처방식을 익힌다.

8. Neural Network의 단점

- 과도한 파라미터 개수
- 많은 훈련 데이터 필요 \Rightarrow 과대적합 위험성 \uparrow

One-Shot-Learning: 이미지 인식 분야에서 사용되는 각 class에 따른 하나의 training 이미지만으로, 분류 모델 생성하는 기법

- 입력 이미지와 원 이미지 간의 유사도를 이용해 구현한다.
- $D(x_1, x_2)$: 두 이미지 차이를 나타내는 distance function을 이용하여 어떤 임계값 τ (타reshold)보다 크면 두 이미지가 다른 것이고, 작으면 비슷하다.
- 인식하고자 하는 이미지가 입력으로 들어오면, 유사도 함수를 사용해 분류할 수 있다.

9. Type Conversion \Rightarrow 자동 형 변환10. $f \rightarrow b \rightarrow t$ (cnt)

```
def Hanoi(f, b, t, cnt):
    if cnt == 0:
        return
    Hanoi(f, t, b, cnt - 1)
    print(f, t)
    Hanoi(b, f, t, cnt - 1)
```