

Bayes Theorem in Machine (Deep) Learning

딥러닝에 어떻게 적용될까?

소프트웨어 공대 강의

노기섭 교수

(kafa46@cju.ac.kr)

What happens in Deep Learning?

■ 딥러닝에서 발생한 상황,

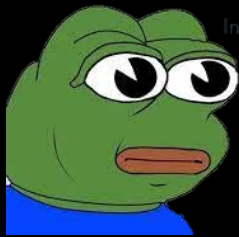
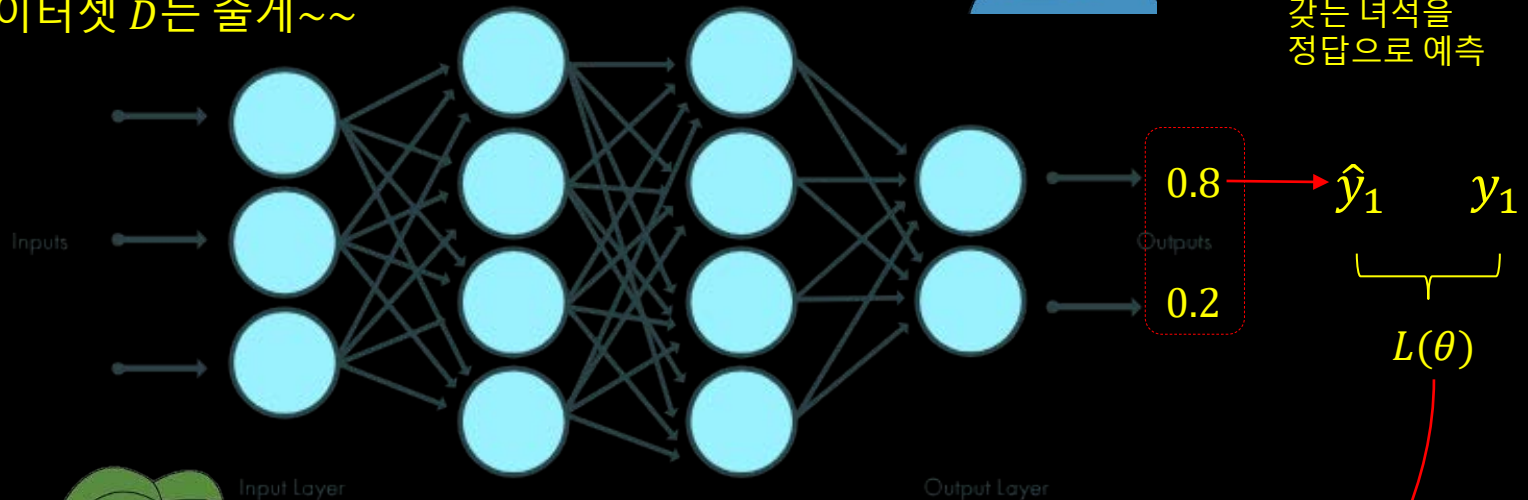
정답과 가장 가까운 출력을 생성하는
파라미터 집합 θ 를 찾아라!

아... 데이터셋 D 는 줄게~~

$y = ax + b$??
무작정 파라미터
 $\theta = \{a, b\}$ 찾으라고??

뭐래! 장난하냐?

가장 큰 확률값
갖는 녀석을
정답으로 예측



교수님~~ 파라미터 θ 찾는거랑
Bayes Theorem하고 무슨 상관이래요?

적절하게 모든
파라미터 θ 를
업데이트

Where is Bayes Theorem in Deep Learning?

- 앞에서 설명한 상황을 Bayes Theorem으로 생각하면?

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

$$\text{Dataset } D = \{ (x_i, y_i) \}_{i=1}^n$$

	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	...	grade	sqft_above	sqft_basement	price	type
n	3	1.00	1180	5650	1.0	0	0	...	7	1180	0	221900.0	아파트
	3	2.25	2570	7242	2.0	0	0	...	7	2170	400	538000.0	빌라
	2	1.00	770	10000	1.0	0	0	...	6	770	0	180000.0	전원주택
	4	3.00	1960	5000	1.0	0	0	...	7	1050	910	604000.0	다가구
	3	2.00	1680	8080	1.0	0	0	...	8	1680	0	510000.0	아파트

$$P(y|x) = P(\text{정답}|\text{데이터})$$

데이터가 주어졌을 때
정답 맞출 확률을 최대화

Dataset D 주어졌을 때 정답 맞출
확률을 최대화하는 파라미터 θ

$$P(\theta|D)$$

Bayes Theorem 적용

$$P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{P(D)}$$

Possible Learning in Bayesian

■ Maximum Likelihood Estimation (MLE)

- Same as frequentist!
- Dataset only!

목표: 오직 Likelihood만 최대화

$$P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{P(D)} \approx P(D|\theta)$$

■ Maximum A Posterior (MAP)

- $P(D)$: 알고(given) 있다고 가정(상수 취급)
 - argmax 과정에서 관계없으므로 생략
- Overfitting에 강함
- Bayesian 관점

목표: Likelihood + Posterior 최대화

$$P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{P(D)} \approx P(D|\theta) \times P(\theta)$$

■ Bayesian Inference (Variation Inference)

- Likelihood, Posterior, Evidence 모두 고려
- Computing $P(D)$ is intractable
- Alternatively, using Variational Inference
- $P(\theta|D)$ 계산이 어렵기 때문에 우리가 알고 있는 함수를 이용하여 잘 모사하도록 접근

$$P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{P(D)} \approx Q(\theta|\theta')$$

목표: P 를 잘 흉내내는 Q 의 파라미터 θ' 찾기



수고하셨습니다 ..^^..