**[베이지안 뉴럴 네트워크 vs 드롭아웃을 적용한 뉴럴 네트워크]**

위 논문에서는 NN 에 드롭아웃을 적용한 것이 Gaussian Process 의 베이지안 추정과 동일하다는 것을 보인다.

모델의 수정없이, 많은 계산이 필요하지 않은 불확실성을 구할 수 있다.

Uncertainty in Deep Learning 이라는 Thesis 에서는 NN에 드롭아웃을 적용한 것이 베이지안 뉴럴 네트워크와 동일하다는 것을 보인다.

가우시안 프로세스와 베이지안 뉴럴 네트워크가 무슨 차이인지는 정확하게는 모르겠지만 수식은 비슷한 것 같다.

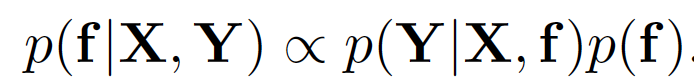
1. 가우시안 프로세스의 베이지안 추정

- 먼저, 가우시안 프로세스의 베이지안 추정의 방법을 살펴보면,

Training dataset: {x1, … , xn}

Output : {y1, … , yn}

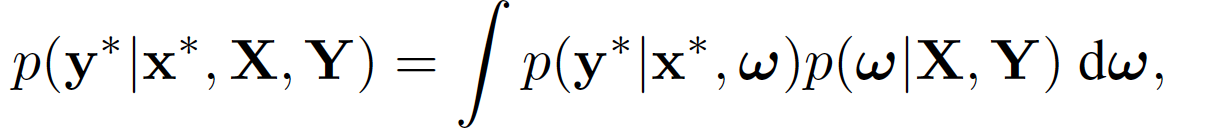
가 되도록 하는 y = f(x) 를 찾아보려 한다.



- 여기서 function 의 분포를 가우시안 프로세스로 모델링 함으로써 posterior 를 찾아보려고 한다.

- 여기서 posterior 를 찾기 위해서 Variational Inference (VI) 를 사용한다.

- 새로운 input 인 x\* 에 대한 predictive distribution 은



데이터를 알고 있을 때, 새로운 x 가 들어왔을 때 그에 대한 아웃풋 y

(이 식에서 p(w|X, Y) 를 계산할 수 없음.)

- 가우시안 프로세스로 부터 얻은 posterior에 비슷한 근사 분포 q(w) 를 정의한다.

- 두개의 분포를 KL-Divergence 를 minimize 하는 쪽으로 q(w) 를 구한다.

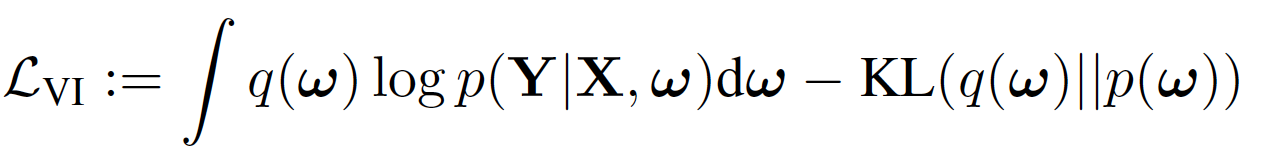
그래서 근사하는 분포 q(w) 를 통한 새로운 x에 대한 분포를 얻을 수 있다.

\텍스트이(가) 표시된 사진

자동 생성된 설명

- (Bishop) 논문에 의해서,

KL Divergence 를 줄이는 것 = log evidence lower bound 를 크게 하는 것



이 식을 크게 하는 q(w) 가 데이터를 잘 설명하고, 실제 분포와 비슷한 분포를 얻을 수 있다.

2. 드롭아웃을 적용한 뉴럴 네트워크

- 드롭아웃을 적용한 뉴럴 네트워크의 식을 잘 정리하면 가우시안 프로세스의 베이지안 추정의 식으로 만들 수 있다.

- 뉴럴 네트워크 최적화 과정의 regularization term 으로 L2 regularization 을 사용한다.

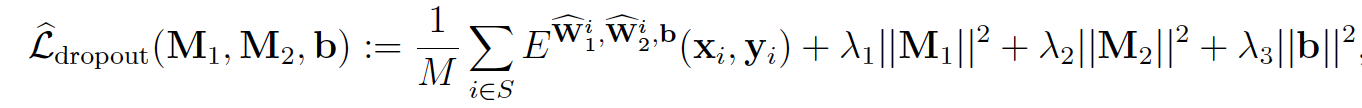
텍스트, 손목시계이(가) 표시된 사진

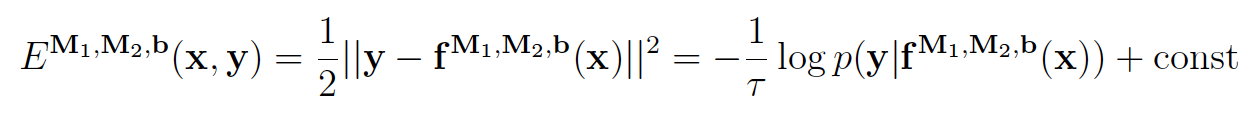
자동 생성된 설명

Loss function

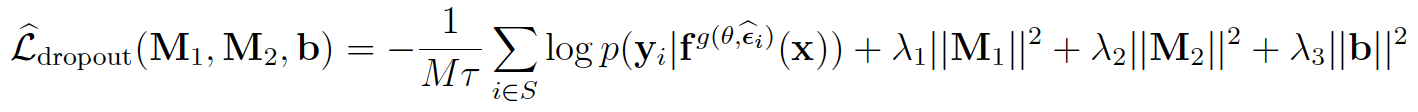
L2 regularization

이후에 Minibatch 거쳐서, 다음과 같은 식이 나온다.



여기서 

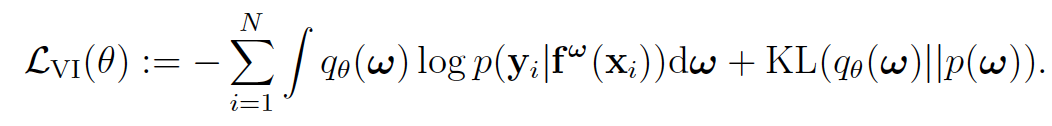
Negative log likelihood 에 상수값을 더한 식으로 바뀌는데, 이 식을 대입하면 아래의 식이 나온다.



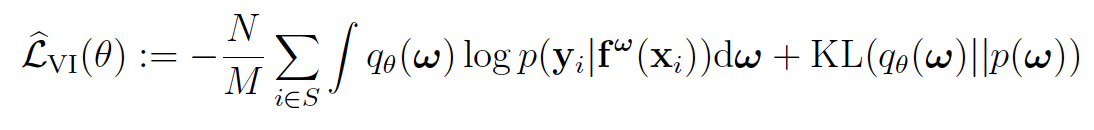
🡪 드롭아웃을 적용한 뉴럴 네트워크의 최적화 과정의 최종 식

3. 베이지안 뉴럴 네트워크

- 베이지안 뉴럴 네트워크의 논문에서는 다음의 식을 maximize 하는 쪽으로 q(w) 를 업데이트한다.



Minibatch



Reparametrization trick

텍스트이(가) 표시된 사진

자동 생성된 설명

여기서 Integral -> Expectation 으로 바꾸고, MC sampling 을 해서,

텍스트이(가) 표시된 사진

자동 생성된 설명

이 식을 posterior 를 계산하기 위해 사용한다.

베이지안의 optimization 최종 식이 된다.

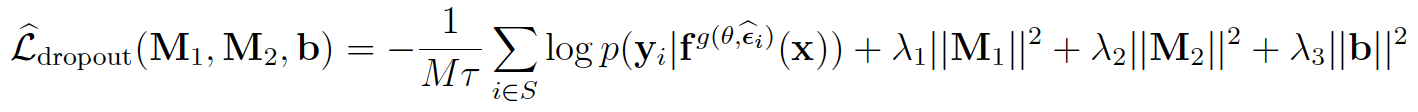
4. 베이지안과 드롭아웃

* 베이지안

텍스트이(가) 표시된 사진

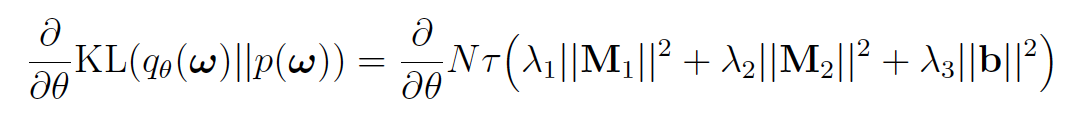
자동 생성된 설명

* 드롭아웃



거의 유사한데, 차이점은 regularization term이

베이지안은 KLD 이고 드롭아웃은 L2 Normalization 이라는 것

여기서 

이렇게 되도록 p(w) 를 잘 맞춰준다면,

텍스트, 시계, 손목시계이(가) 표시된 사진

자동 생성된 설명

베이지안과 드롭아웃을 이용한 방법의 optimization 식이 동일해진다.