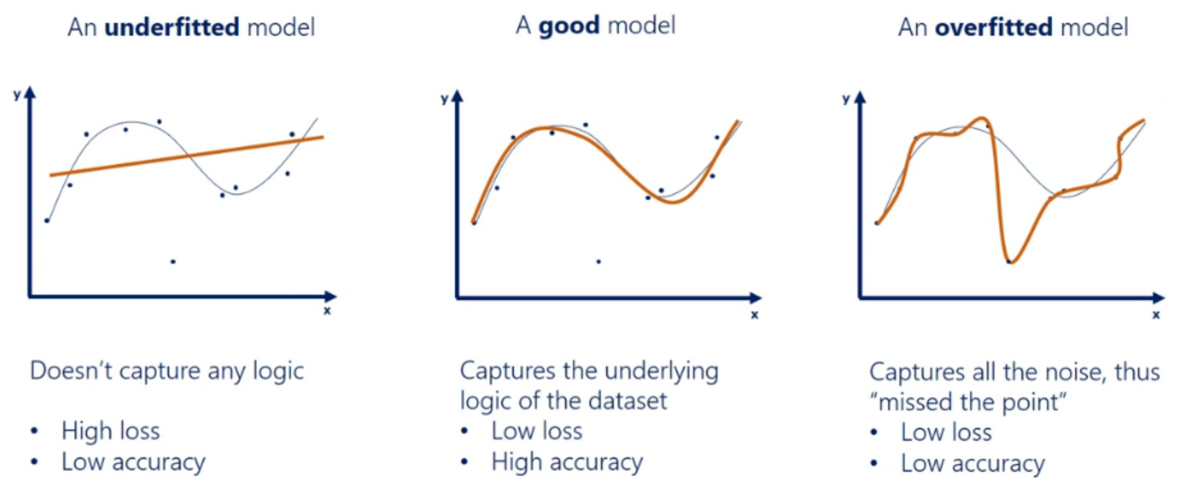
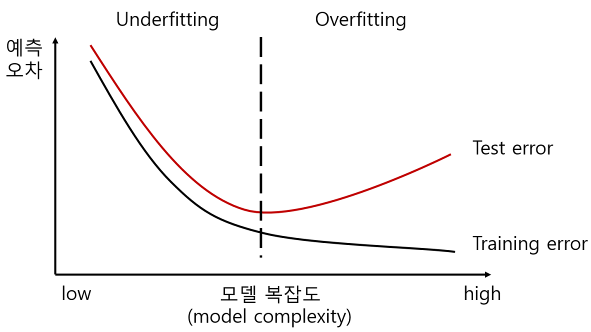
Bayesian (3)

Overfitting





오버피팅을 해결하는 법

- 학습 데이터 증폭
- regularization: 오버피팅이 일어날 시 가중치가 커지는 현상을 줄여주기 위헤 penalty 도입
- 베이지안 접근: 적용시 오버피팅이 일어나지 않음 (Maximum Likelihood 방법을 통한 모수의 추정은 모형의 오버피팅 문제를 발생시킴)

EM Algorithm

Bayesian Inference (베이즈 추론)

: 이전의 경험과 현재의 증거를 토대로 어떤 사건의 확률을 추론하고자 베이즈 정리를 사용하는 것

베이즈 정리: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

$P(A)$: 사전확률

$P(B)$: 증거

$P(B|A)$: 가능도

$P(A|B)$: 사후확률; B라는 증거가 관찰된 후의 명제에 대한 확률. (증거를 보고 변화된 믿음의 정도)

Variational Inference (변분 베이즈 방법)

: 모수분포를 **최적화**를 통해서 찾아내는 방법

- 베이지안 통계학에서 **샘플링**을 통해 모수 분포 찾아내는 방법: acceptance-rejection method, metropolitan-hasting algorithm, gibbs sampler

특징

- 최적화 기반 방법으로 목표 분포 찾아냄
- local optima에 빠질 수 있음
- low computing cost

MCMC 특징

- 샘플링 기반 방법으로 목표 분포 찾아냄
- global optima 보장
- high computing cost

<목적 함수 정의>

다루기 어렵지만 true인 p함수와, 다루기 쉽지만 true는 아닌 q함수 근사화. 두 함수의 거리를 **Kullback-Leibler (KL) Divergence** 방법으로 정의하고 최소화함.

$$KL(q(z) \mid\mid p(z \mid \mathbf{x})) \equiv \int q(z) log \frac{q(z)}{p(z \mid \mathbf{x})} dz$$

KL 정의를 사용하면 다음 수식이 만족함

$$log\ p(\mathbf{x}) = E_q(log\ p(z, \mathbf{x})) - log\ q(z)) + KL(q(z) \mid\mid p(z \mid \mathbf{x}))$$

- $E_q(log\ p(z, \mathbf{x})) - log\ q(z))$을 목적함수 $F(q)$라고 두고 최대화 시키면 KL 거리가 줄어듦. 이때 $F(q)$= Evidence Lower Bound (ELBO)임 (왜냐면 $log\ p(\mathbf{x})$가 베이즈 룰에서 evidence꼴임)

<최적화>

- ELBO 최적화시 미분해서 0이 나오는 점을 찾는게 목표. 이때 ELBO는 Functional, 즉 함수 q를 인풋으로 받기 때문에 일반적인 미분법을 따르지 않음.
- 수식을 유도하면… ELBO를 최대화시키는 q^*를 찾아야하는데 이때 **Mean Field Assumption**을 사용함

$$q(z) = \prod q_j(z_j)$$

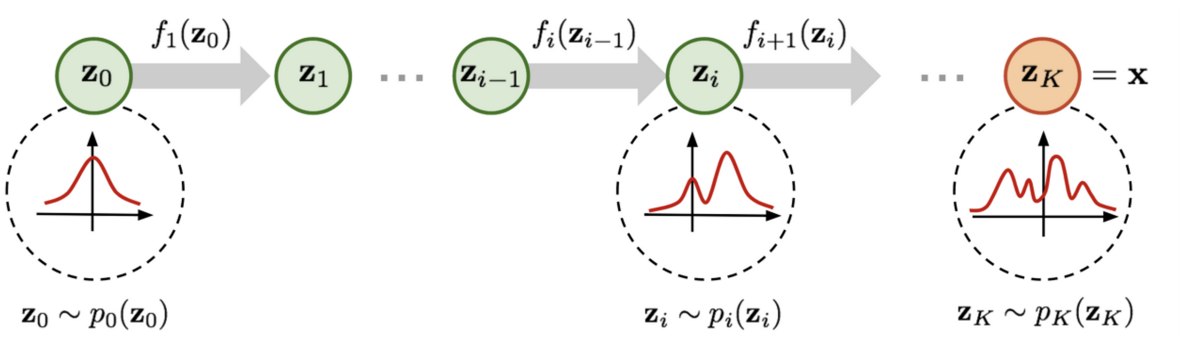
- 수식 유도하면 $q_j^*(z_j) \propto exp(E_{q_{-j}}[log\ p(z_j \mid z_{-j}, \mathbf{x})])$

Reparameterization Trick

: VAE에서 back propagation이 가능하도록 하는 trick

- VAE: AutoEncoder의 확률모델적 변형. 새로운 데이터를 샘플링할 수 있게 해줌

Normalizing Flow



: 어떤 쉬운 분포 z에 연산 f들을 해주어 복잡한 확률분포 $p(x)$를 모델링할 수 있도록 하는 것.

- 직접적으로 모델의 입력 데이터에 대한 likelihood 계산함
- flow based generative model