

과제 A

1. SGVB에서 Reparametrization trick을 쓰는 이유는?

Parameter θ 를 Gaussian에서 샘플링하는데, 이때 미분가능한 파이프라인을 만들기 위하여 사용.

$\theta \sim N(\mu, \sigma^2)$ 에서 바로 샘플링하는 대신, μ 와 σ 를 모델링하고 $N(0, 1)$ 에서 샘플링한 값과 결합하여 샘플링.

2. Sparse VD에서 Local reparameterization을 사용하는 이유는?

Central limit theorem을 이용하여 minibatch size를 늘리면 ELBO 평균의 variance를 줄일 수 있음. 하지만 covariance를 줄이는 것은 불가능하기 때문에, 이를 해결하려면 minibatch안의 데이터간 covariance를 0으로 만들기 위해서 도입.

Minibatch 안의 데이터 수만큼 θ 를 reparameterization을 이용해서 sampling하면, minibatch각각에 적용되는 θ 는 서로 독립되어 ELBO간의 covariance를 0으로 만드는 것이 가능.

하지만 이경우 연산 코스트가 비싸지므로, 이전 레이어에서 넘어온 Activation map A 가 Gaussian이고 역시 Gaussian θ 를 곱한 결과인 B 도 Gaussian임을 이용해서 B 에서 γ 평균과 δ variation를 $\zeta \sim N(0,1)$ 로 reparameterization으로 sampling하였다.

3. Sparse VD에서 weight의 sparsity가 유도되는 이유는?

Weight 별로 dropout α ($1-p/p$)를 학습하고, α 가 제한없이 커질 수 있게 학습하여서 $p=1$ 이 되는 weight들을 학습하여 매우 sparse한 학습이 가능하게 됨.

4. Variational Dropout의 prior가 log uniform인 이유는?

Sparsity induce prior로써 Zero 근처에서 높은 density를 가지고 ELBO에서 Expected loglikelihood term에 대해서만 학습 될 수 있도록 하는 조건을 만족함.