CS 3750 Machine Learning
Lecture 6

Variational approximations

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square

Variational approximation

Assume we have a function $f(Z)$ that is hard to calculate

Example: *posterior probability in a complex BBNs*

$$P(Z \mid X)$$

- *this inference can be very hard*

Idea: replace calculations of $f(Z)$ with an optimization over a simpler parametric function $q(Z \mid \lambda)$

$$f(Z) \sim \max_{\lambda} q(Z \mid \lambda)$$

Variational lower bound

Let *X denote observed variables and*
Z denote target variables

$$P(Z \mid X) = \frac{P(X,Z)}{P(X)}$$

$$\log P(Z \mid X) = \log P(X,Z) - \log P(X)$$

$$\log P(X) = \log P(X,Z) - \log P(Z \mid X)$$

Assume some distribution: $Q_\theta(Z \mid X)$ *defined by parameters* θ

Average both sides with E_{Q_θ}

$$\sum_Z Q_\theta(Z \mid X) \log P(X) = \sum_Z Q_\theta(Z \mid X) \log P(X,Z) - \sum_Z Q_\theta(Z \mid X) \log P(Z \mid X)$$

$$\log P(X) = E_{Q_\theta}\left(\log P(X,Z)\right) - E_{Q_\theta}\left(\log P(Z \mid X)\right)$$

Variational lower bound

$$\log P(X) = E_{Q_\theta}\left(\log P(X,Z)\right) - E_{Q_\theta}\left(\log P(Z \mid X)\right)$$

$$\log P(X) = \sum_Z Q_\theta(Z \mid X) \log P(X,Z) - \sum_Z Q_\theta(Z \mid X) \log P(Z \mid X)$$

$$\log P(X) = \sum_Z Q_\theta(Z \mid X) \log P(X,Z) - \sum_Z Q_\theta(Z \mid X) \log P(Z \mid X) + \sum_Z Q_\theta(Z \mid X) \log Q_\theta(Z \mid X) - \sum_Z Q_\theta(Z \mid X) \log Q_\theta(Z \mid X)$$

Kullback-Leibler divergence: distance between 2 distributions

$$KL(Q \mid P) = \sum_Z Q_\theta(Z \mid X) \log Q_\theta(Z \mid X) - \sum_Z Q_\theta(Z \mid X) \log P(Z \mid X)$$

Functional:

$$F(Q,P) = \sum_Z Q_\theta(Z \mid X) \log P(X,Z) - \sum_Z Q_\theta(Z \mid X) \log Q_\theta(Z \mid X)$$

$$\log P(X) = F(Q,P) + KL(Q \mid P)$$

Variational lower bound

$$\log P(X) = F(Q,P) + KL(Q\,|\,P)$$

distance between $Q_\theta(Z\,|\,X), P(Z\,|\,X)$

Always ≥ 0

Equals 0 if $Q_\theta(Z\,|\,X) = P(Z\,|\,X)$

We can optimize the approximation $Q_\theta(Z\,|\,X)$ by minimizing

$$\min_\theta KL(Q_\theta\,|\,P)$$

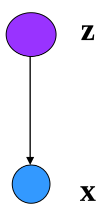
We can also do this by maximizing $F(Q_\theta, P)$

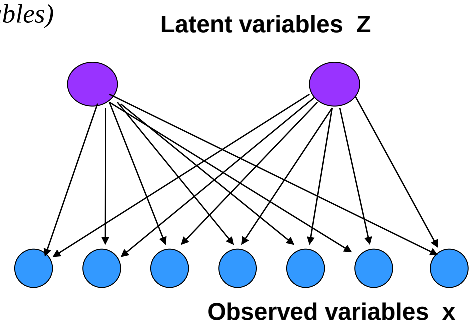
$$\max_\theta F(Q_\theta, P)$$ ← Often much easier

$$F(Q,P) = \sum_Z Q_\theta(Z\,|\,X) \log P(X,Z) - \sum_Z Q_\theta(Z\,|\,X) \log Q_\theta(Z\,|\,X)$$

Latent variable models

Let *X denote observed variables and*
Z denote hidden (latent variables)





Inference opposite the links is hard: $P(Z\,|\,X)$

Solution: Define a simpler distribution : $Q_\theta(Z\,|\,X)$ to approximate $P(Z\,|\,X)$

Optimize: $\max_\theta F(Q_\theta, P)$

$$F(Q,P) = \sum_Z Q_\theta(Z\,|\,X) \log P(X\,|\,Z) + \sum_Z Q_\theta(Z\,|\,X) \log P(Z) - \sum_Z Q_\theta(Z\,|\,X) \log Q_\theta(Z\,|\,X)$$