Neural Computer

Thesis Subtitle

Renan Monteiro Barbosa

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Turpis egestas pretium aenean pharetra magna ac placerat vestibulum.

Implementing some equations that repreent the Brain as a Dynamical System:
Latent factor $Z->X[TSC]$ Observation
The distribution of x is compatible with the sampled Z
$P(x|z)$ – P of x given z conditional probablity
This is Bayesian
The joint probability of x and z occuring togehter equals the probability of Z and x given z
$P(x,z) = P(z) \cdot P(x|z)$
It is important to note how we can parametrize this probability by leveragin a distribution and rely on the mean field theory.
$P(x) = \frac{1}{\sigma\sqrt{2\pi}}$

Isotropic Gaussian

minimize the KL divergence
$D_{KL}[P(x)||P_\theta(x)] = \sum_{x}^{states} P(x) \cdot \log \frac{P(x)}{P_\theta(x)}$
maximize the expected log probability
$\sum_{x}^{states} P(x) \cdot \log P_\theta(x)$

Importance Sampling

Importance sampling is a variance reduction technique that can be used in the Monte Carlo method.
Example using the trees
1) Measure 50-50 from both regions
2) Corect for non-uniform Sampling

Variational Inference

There is a network which is trained to learn the Variational distribution
variational distribution: $Q_\theta(z|x)$
This is also referenced in the Free energy as the recognition model
$P_\theta(x) = \sum_z P_\theta(x|z) \frac{P_\theta(z)}{Q_\theta(z|x)} Q_\theta(z|x)$
Sampling Correction $\frac{P_\theta(z)}{Q_\theta(z|x)}$

ELBO - Evidence Lower Bound

In variational Bayesian methods, the evidence lower bound (often abbreviated ELBO, also sometimes called the variational lower bound[1] or negative variational free energy) is a useful lower bound on the log-likelihood of some observed data.

Minimizig the KL = Maximizing the ELBO

Variational Free Energy $:= \mathbb{E}[\log q(z)] - \mathbb{E}[\log p(x|z)p(z)]$

ELBO $:= \mathbb{E}[\log p(x|z)p(z)] - \mathbb{E}[\log q(z)]$

Math Concepts

Expectation of a random variable

$\mathbb{E}[f(x)] = \int x f(x)\, dx$

Chain rule of probability

$P(x, y) = P(x|y)P(y)$

Bayes' Theorem

$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$

Kullback-Leibler Divergence

The **Kullback-Leibler Divergence** is a measure of the difference between two probability distributions, quantifying how much one distribution diverges from another.

$D_{KL}(P \parallel Q) = \int p(x) log \left(\frac{p(x)}{q(x)}\right)\, dx$

Properties:

- Not symmetric
- Always ≥ 0
- It is equal to 0 if and only if P = Q

Deriving the ELBO

This is the log likelihood of our data:
$\log p_\theta(x) = \log p_\theta(x)$

Now multiply by 1: $\log p_\theta(x) = \log p_\theta(x) \int q_\phi(z|x)\, dz$

Bring inside the integral: $\log p_\theta(x) = \int \log p_\theta(x) q_\phi(z|x)\, dz$

Definition of Expectation: $\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x)]$

Apply the equation $p_\theta(x) = \frac{p_\theta(x,z)}{p_\theta(z|x)}$

$\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)} \left[\frac{p_\theta(x,z)}{p_\theta(z|x)}\right]$

Multiply by 1

$\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)} \left[\frac{p_\theta(x,z)}{p_\theta(z|x)} \cdot \frac{q_\phi(z|x)}{q_\phi(z|x)}\right]$

Split the Expectation

$\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)} \left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\right] + \mathbb{E}_{q_\phi(z|x)} \left[\frac{q_\phi(z|x)}{p_\theta(z|x)}\right]$

Definition of KL divergence:

$\log p_\theta(x) = \mathbb{E}_{q_\phi(z|x)} \left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\right] + D_{KL}\left(q_\phi(z|x) \parallel p_\theta(z|x)\right)$

Remember that $D_{KL} \geq 0$

$$\log p_\theta(x) = \underbrace{\mathbb{E}_{q_\phi(z|x)}\left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\right]}_{\text{ELBO}} + \underbrace{D_{KL}\left(q_\phi(z|x) \parallel p_\theta(z|x)\right)}_{\geq 0}$$

We can deduce that

$\log p_\theta(x) \geq \mathbb{E}_{q_\phi(z|x)}\left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\right]$

because the ELBO is the lower bound log likelihood of our data, so if we want to maximize the log likelihood we will be maximizing the ELBO at the same time.

Maximizing the ELBO means:

- Maximizing the first term: Maximizing the reconstruction likelihood of the decoder
- Minimizing the second term: Minimizing the distance between the leanred dsitribution and the prior belief we have over the latent variable.