Mean Field Approximation

Munish(190519) and Sai Ashwith(190284)

7th February 2022

1 Variational Inference

If we have a density, say $p^*(x)$, sampling from which is not computationally viable, then we try to approximate $p^*(x)$, with a distribution which can be sampled easily.

We consider a family of distribution say, $q_i(x)$, which is easy to sample and using KL Divergence as our distance metric, we find $q^*(x) \in q_i(x)$, such that it's distance is minimum to target $p^*(x)$ among the family $q_i(x)$.

$$q^*(x) = \underset{q_i(x)}{\arg\min}\ KL\{q_i(x) \| p^*(x)\}$$

2 Mean Field Approximation

For Mean Field Approximation, we consider a family of distributions $\mathbf{Q}$, where the individual dimensions z_i are independent of each other.

$$\mathbf{Q} \;=\; \left\{ q_{\mathbf{z}}(\mathbf{z}) \mid \; q_{\mathbf{z}}(\mathbf{z}) = \prod_{i=1}^{n} q_{\mathbf{z}_i}(z_i) \right\} \tag{1}$$

And we aim to find a distribution $q^*(\mathbf{z})$ from this family which has minimum KL Divergence with the target distribution.

$$q^*_{\mathbf{z}}(\mathbf{z}) \;=\; \underset{q_{\mathbf{z}} \,\in\, \mathbf{Q}}{\arg\min}\ KL\left[\, q_{\mathbf{z}}(\mathbf{z}) \,||\, p^*_{\mathbf{z}}(\mathbf{z})\right] \tag{2}$$

And KL Divergence is defined as,

$$KL\left[q(\mathbf{z})\,||\,p(\mathbf{z})\right] \;=\; \int\limits_{\mathbf{z}} q(\mathbf{z})log\left(\frac{q(\mathbf{z})}{p(\mathbf{z})}\right)d\mathbf{z}$$

To obtain general update form for Mean Field Approximation, we optimise above equation for a given $p^*_{\mathbf{Z}}(\mathbf{z})$. In general for higher dimensions, Gradient Descent is not computationally feasible and in this case, each z_i is independent of each other, so we can use Coordinate Descent Algorithm.

Some notations which are used later:

$$q_i = q_i(z_i),\; q_{-k} = \prod_{i \neq k} q_i(z(i))$$

Coordinate Descent Algorithm :

Step 1 : Compute

$$q_1 = \arg\min_{q_1} KL[q\,||p^*];\; \textit{update}\ \mathbf{q}$$

Step 2 : Compute

$$q_2 = \arg\min_{q_2} KL[q\,||p^*];\; \textit{update}\ \mathbf{q}$$

$\vdots$

Step N : Compute

$$q_N = \arg\min_{q_N} KL[q\,||p^*];\; \textit{update}\ \mathbf{q}$$

Step N+1 : Check Convergence, if converged then **exit**, else go back to **step 1**

Derivation : Mean Field Approximation

For any given $p^*_{\mathbf{Z}}(\mathbf{z})$, consider updating q_k, $1\leq$ k $\leq$N, then the objective function (KL-divergence), viewed as a function of z_k is,

$$\int\limits_{\mathbf{z}} \prod_{i=1}^{n} q_i(z_i) \log\left(\frac{\prod\limits_{j=1}^{n} q_j(z_i)}{p^*(\mathbf{z})}\right) \prod_{l=1}^{n} dz_l$$

$$= \int\limits_{\mathbf{z}} q_k \prod_{i\neq k} q_i \;(\log(q_k) + \log(\prod_{i\neq k} q_i) - \log(p^*(\mathbf{z}))dz_k \prod_{j\neq k} dz_j$$

$$= \int\limits_{z_k} q_k \log(q_k) dz_k \int\limits_{z_{-k}} \prod_{i\neq k} q_i dz_i + \int\limits_{z_k} q_k dz_k \int\limits_{z_{-k}} \prod_{i\neq k} q_i \log(\prod_{l\neq k} q_l) \prod_{j\neq k} dz_j -$$

$$\int\limits_{z_k} q_k \Big(\int\limits_{z_{-k}} log(p^*(\mathbf{z})) \prod_{i\neq k} q_i \; dz_i \Big) dz_k$$

We know that,

$$\int\limits_{z_p} q_p(z_p) dz_p \; = \; 1 \,, \quad for \, 1 \leq p \leq N$$

So,

$$\int\limits_{z_{-k}} log(p^*(\mathbf{z})) \prod_{i\neq k} q_i \; dz_i = \mathop{\mathbb{E}}_{\prod_{\mathrm{i\neq k}} \mathrm{q_i}} \left[\log(\mathrm{p}^*(\mathbf{z})) \right]$$

And the term $\displaystyle\int\limits_{z_{-k}} \prod_{i \neq k} q_i \; \log(\prod_{l \neq k} q_l) \prod_{j \neq k} dz_j$ doesn't depend on q_k, so it's just a constant for our optimization problem. Therefore, the expression that is to be optimized boils down to:

$$\int\limits_{z_k} q_k \log(q_k) - \int\limits_{z_k} q_k \underset{\prod_{\mathrm{i \neq k}} \mathrm{q_i}}{\mathbb{E}} \Big[\log(\mathrm{p}^*(\mathbf{z})) \Big] \, dz_k + const.$$

$$= \int\limits_{z_k} q_k \log \left(\frac{q_k}{exp\Big(\mathbb{E}_{\prod_{\mathrm{i \neq k}} \mathrm{q_i}} \big[\log(\mathrm{p}^*(\mathbf{z})) \big] \Big)} \right) dz_k + const. \tag{3}$$

Equation (3) is nothing but $\mathrm{KL}\Bigg[q_k \Bigg|\Bigg| \mathbb{E}_{\prod_{\mathrm{i \neq k}} \mathrm{q_i}} \big[\log(\mathrm{p}^*(\mathbf{z})) \big] \Bigg]$.

Therefore, to minimize equation (3), we take $q_k = \mathbb{E}_{\prod_{\mathrm{i \neq k}} \mathrm{q_i}} \big[\log(\mathrm{p}^*(\mathbf{z})) \big]$.

This is **Mean Field Approximation**.

Mean Field Approximation applied on Ising Model

Ising Model, described in previous lecture, is given by,

$$p_{\mathbf{Y}}(\mathbf{y}) = exp(\beta \sum_{<i,j>} y_i \, y_j + \sum_i b_i y_i) \tag{4}$$

where, β and b_i are positive constants and $< i, j >$ denotes that y_i and y_j are neighbours.
Mean Field approximation be

$$q_{\mathbf{z}}(\mathbf{z}) = \prod_i q_i(z_i)$$

Then the update equation for k, 1$\leq$ k $\leq$ is

$$\log(q_k) = \mathop{\mathbb{E}}_{\text{q-k}}[\log(p_{\mathbf{Y}}(\mathbf{y}))]$$

$$\log(q_k) = \mathop{\mathbb{E}}_{\text{q-k}}[\beta \sum_{<i,j>} y_i\, y_j + \sum_i b_i y_i]$$

Now, all terms that don't contain y_k can be clubbed under a constant, which is independent of y_k and all y_i not neighbours of y_k are independent of y_k. So,

$$\log(q_k) = \mathop{\mathbb{E}}_{\text{q-k}}[\beta \sum_{<k,j>} y_k\, y_j + b_k y_k] + const$$

Also, Expectation is w.r.t. all y_i except k.

$$\log(q_k) = \beta\, y_k \left(\sum_{j,<k,j>} \mathop{\mathbb{E}}_{q_j}[y_j] \right) + b_k\, y_k + const$$

The sum is over all neighbours of k.
Denoting $\mathbb{E}_{q_j}[y_j]$ as μ_j, since it is mean of y_j w.r.t. q_j.

$$\log(q_k) = y_k \left(\beta \sum_{j \in ngb(k)} \mu_j + b_k \right) + const$$

where $ngb(k)$ is the set of neighbours of k.
Say, $M_k = \beta \sum_{j \in ngb(k)} \mu_j + b_k$, which is independent of the value of y_k. Then,

$$q_k = C.exp(y_k\, M_k)$$

Since we know y_k only takes $\{-1, +1\}$ values and q_k must sum to 1 over this set, we can calculate constant C, which is

$$C = \frac{1}{e^{M_k} + e^{-M_k}}$$

So,

$$q_k = \frac{e^{y_k\, M_k}}{e^{M_k} + e^{-M_k}}$$

Also, we can now find out the value of μ_k as defined above,

$$\mu_k = \mathop{\mathbb{E}}_{q_k}[y_k] = 1Ce^{M_k} + (-1)Ce^{-M_k} = \frac{e^{M_k} - e^{-M_k}}{e^{M_k} + e^{-M_k}} = tanh(M_k).$$