Problem Set 5

Nima Leclerc (nleclerc@seas.upenn.edu, PennID = nleclerc)
Collaborated with Abhinav R.

ESE 546 (Principles of Deep Learning)
School of Engineering and Applied Science
University of Pennsylvania
Total time = 10

December 14, 2020

Problem 1:
Solution:

Provided the distributions $q(w)$ and $p(w)$, we have that $p(w)$ is given by

$$p(w) = \frac{e^{-\beta\Phi(w)}}{Z(\beta)}$$

the quantity $(q||p)$ can be used to solve the variational inference problem. Hence, we can first expand out this KL-divergence.

$$KL(q||p) = \sum_{w\in W} q(w) \log \frac{q(w)}{p(w)} = \sum_{w\in W} q(w) \log \frac{q(w)}{\frac{e^{-\beta\Phi(w)}}{Z(\beta)}}$$

$$= \sum_{w\in W} q(w)[\log q(w) - (\log e^{-\beta\Phi(w)} - \log Z)]$$

$$= \sum_{w\in W} (q(w)\log q(w) + q(w)\log Z) + \beta \sum_{w\in W} q(w)\Phi(w)$$

$$= \beta \sum_{w\in W} q(w)\Phi(w) + \sum_{w\in W} q(w)\log q(w) + \log Z(\beta) \sum_{w\in W} q(w)$$

$$= \beta \sum_{w \in W} q(w)\Phi(w) + \sum_{w \in W} q(w) \log q(w) + \log Z(\beta)$$

Hence, we have arrived at our final expression for $KL(q||p)$.

(b) We can now show that,

$$\mathbb{E}_{w \sim p(w)}[\Phi(w)] = -\frac{\partial \log Z(\beta)}{\partial \beta}$$

We can write,

$$\mathbb{E}_{w \sim p(w)}[\Phi(w)] = \int_{w \in \mathcal{W}} \Phi(w) p(w) dw$$

$$= \int_{w \in \mathcal{W}} \Phi(w) \frac{e^{-\beta\Phi(w)}}{Z(\beta)} dw = \frac{1}{Z} \int_{w \in \mathcal{W}} \Phi(w) e^{-\beta\Phi(w)} dw$$

We can now show that $-\frac{\partial \log Z(\beta)}{\partial \beta}$ gives this expression above. We have that,

$$Z(\beta) = \int_{w \in \mathcal{W}} e^{-\beta\Phi(w)}$$

Hence,

$$\frac{\partial \log Z(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \log \int_{w \in \mathcal{W}} e^{-\beta\Phi(w)} dw$$

$$= (\int_{w \in \mathcal{W}} e^{-\beta\Phi(w)} dw)^{-1} \int_{w \in \mathcal{W}} \frac{\partial}{\partial \beta} e^{-\beta\Phi(w)} dw$$

$$= -\frac{1}{Z} \int_{w \in \mathcal{W}} \Phi(w) e^{-\beta\Phi(w)} dw$$

Hence, we get that

$$\mathbb{E}_{w \sim p(w)}[\Phi(w)] = \frac{1}{Z} \int_{w \in \mathcal{W}} \Phi(w) e^{-\beta\Phi(w)} dw = -\frac{\partial \log Z(\beta)}{\partial \beta}$$

So,

$$\mathbb{E}_{w \sim p(w)}[\Phi(w)] = -\frac{\partial \log Z(\beta)}{\partial \beta}$$

(c) Given that,

$$\mathcal{Q}_2 = \{q(w) : q(w) = \prod_{i=1}^{N} q(w_i)\}$$

with data generated from,

$$p(w) \propto 1 - \prod_{i=1}^{N} w_i$$

we can now find the minimizer,

$$q^{\star} = \min_{q \in \mathcal{Q}_2} KL(q||p)$$

We can first write out our expression for $KL(q||p)$.

$$KL(q||p) = \sum_{w \in W} q(w) \log \frac{q(w)}{p(w)} = \sum_{w \in W} \prod_{i=1}^{N} q(w_i) \log \frac{\prod_{i=1}^{N} q(w_i)}{p(w)}$$

$$= \prod_{i=1}^{N} q(w_i = 1) \log \frac{\prod_{i=1}^{N} q(w_i = 1)}{p(w = 1)} + \prod_{i=1}^{N} q(w_i = 0) \log \frac{\prod_{i=1}^{N} q(w_i = 0)}{p(w = 0)}$$

$$= q(w = 1) \log \frac{q(w = 1)}{p(w = 1)} + \sum_{w=0} [\prod_{i=1}^{m} q(w_i = 0) \prod_{j=1}^{N-m} q(w_j = 0) \log \frac{\prod_{i=1}^{m} q_i(w_i = 0) \prod_{j=1}^{N-m} q(w_j = 0)}{p(w = 1)}]$$

$$= q(w = 1) \log \frac{q(w = 1)}{p(w = 1)}$$

$$+(2^N - 1)[\prod_{i=1}^{m} q(w_i = 0) \prod_{j=1}^{N-m} q(w_j = 1) \log \frac{\prod_{i=1}^{m} q_i(w_i = 1) \prod_{j=1}^{N-m} q(w_j = 0)}{p(w = 1)}]$$

We have the necessary condition that $q_i(1) = 0$ and $q_i(0) = 1$. Hence, the above expression for the KL-divergence goes to 0 because of this necessary condition. The above reduces to zero,

$$q^{\star} = \prod_{i=1}^{N} (1 - w_i)$$

Hence, we do find that $q^{\star}(w)$ does not have a form similar to $p(w)$. While the distributions look similar, they are indeed different distributions.

(d) Given that,

$$\mathcal{Q}_1 = \{q(w) : q(w) = \prod_{i=1}^{N} q_i(w)\}$$

we would like to solve the problem,

$$q^\star = \min_{q \in \mathcal{Q}_1} KL(q||p)$$

We can start with the general expression that was obtained in Bishop's textbook for $q_j^\star$.

$$\ln q_j^\star(w_j) = \mathbb{E}_{i \neq j}[\ln p]$$

We now compute the right-hand side provided that we know p.

$$\mathbb{E}_{i \neq j}[\ln p] = \ln(\frac{1}{N}\sum_{j=1}^{N} p(w_j))$$

$$= \ln(\frac{1}{N}\sum_{j=1}^{N}(1 - \frac{\prod_{i \neq j=1}^{N} w_i}{w_j}))$$

$$= \ln(\frac{N}{N} - \frac{1}{N}\sum_{j=1}^{N}(1 - \frac{\prod_{i \neq j=1}^{N} w_i}{w_j}))$$

$$= \ln(1 - \frac{1}{N}\sum_{j=1}^{N}(\frac{\prod_{i \neq j=1}^{N} w_i}{w_j}))$$

From our previous relation, we can exponentiate both sides to get $q_j^\star(w_j)$. This gives,

$$q_j^\star(w_j) = 1 - \frac{1}{N}\sum_{j=1}^{N}\frac{\prod_{i \neq j=1}^{N} w_i}{w_j}$$

Hence, the solution for $q^\star$ is given by

$$q^\star = \prod_{j=1}^{N} q_j^\star = \prod_{j=1}^{N}(1 - \frac{1}{N}\sum_{j=1}^{N}\frac{\prod_{i \neq j=1}^{N} w_i}{w_j})$$

Applying the proper normalization gives,

$$q^\star = \frac{1}{2^N - 1}\prod_{j=1}^{N}(1 - \frac{1}{N}\sum_{j=1}^{N}\frac{\prod_{i \neq j=1}^{N} w_i}{w_j})$$

Problem 2:
Solution:

(a) Refer to attached Jupyter notebook.
(b) Refer to attached Jupyter notebook.
(c) Refer to attached Jupyter notebook.
(d) Refer to attached Jupyter notebook.