Homework Assignment # 4

Due: Friday, April 4, 2025, 11:59 p.m.
Total marks: 100

Question 1. [15 MARKS]

In machine learning, we sometimes want to use an expectation as a loss $L(\theta)$ and compute

$$\nabla_\theta L(\theta) = \nabla_\theta \mathbb{E}_{\boldsymbol{X}\sim P_\theta(\cdot)}[f_\theta(\boldsymbol{X})].$$

We used just such a loss for generative models, where the goal was to learn a parameterized generative model P_θ that generates plausible $\boldsymbol{X}$. However, sampling this gradient is less straightforward than it was for our predictive models (our generalized linear models and extensions use fixed basis and neural networks). You will reason about this issue in this question.

You can use the fact that

$$\nabla_\theta \mathbb{E}_{\boldsymbol{X}\sim P_\theta(\cdot)}[f_\theta(\boldsymbol{X})] = \mathbb{E}_{\boldsymbol{X}\sim P_\theta(\cdot)}[f_\theta(\boldsymbol{X})\nabla_\theta \log P_\theta(\boldsymbol{X})] + \mathbb{E}_{\boldsymbol{X}\sim P_\theta(\cdot)}[\nabla_\theta f_\theta(\boldsymbol{X})] \tag{1}$$

(a) [5 MARKS] Intuitively, it feels like we should be able to sample $\boldsymbol{X}_i \sim P_\theta$, n times, and then use $\frac{1}{n}\sum_{i=1}^n \nabla_\theta f_\theta(\boldsymbol{X}_i)$ to get an estimate of the gradient in (1). Explain why $\frac{1}{n}\sum_{i=1}^n \nabla_\theta f_\theta(\boldsymbol{X}_i)$ is not an unbiased estimate of $\nabla_\theta L(\theta)$. *Hint:* This question is meant to be straightforward. Reason about what $\frac{1}{n}\sum_{i=1}^n \nabla_\theta f_\theta(\boldsymbol{X}_i)$ estimates (its true expectation) and compare that with Equation (1).

(b) [10 MARKS] Some $P_\theta(\cdot)$ allow us to rewrite the expectation in a convenient way. For example, if $\theta = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ and $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we have $\boldsymbol{X} = X_\theta(\boldsymbol{\epsilon}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then,

$$\mathbb{E}_{\boldsymbol{X}\sim P_\theta(\cdot)}[f_\theta(\boldsymbol{X})] = \mathbb{E}_{\boldsymbol{X}\sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})}[f_\theta(\boldsymbol{X})] = \mathbb{E}_{\boldsymbol{\epsilon}\sim \mathcal{N}(\mathbf{0},\mathbf{I})}[f_\theta(X_\theta(\boldsymbol{\epsilon}))] = \mathbb{E}_{\boldsymbol{\epsilon}\sim \mathcal{N}(\mathbf{0},\mathbf{I})}[f_\theta(\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon})]$$

Now, equivalently, we can get an unbiased estimate of $\mathbb{E}_{\boldsymbol{\epsilon}\sim \mathcal{N}(\mathbf{0},\mathbf{I})}[f_\theta(\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon})]$ using a sample average over n sampled $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Explain the procedure to do so.

Coding Questions

For these questions we have provided a Python helper scipt `assign4.py`. In order to run the code, you will need to install the following packages: `numpy`, `torch`, `torchvision`, and `matplotlib`. You can install these packages using `pip` by running

```
pip install numpy torch torchvision matplotlib.
```

To run the script, simply open up the terminal and run `python assign4.py`. This will run the script and generate the necessary plots. You can also run the script in a Jupyter notebook if you prefer.

Question 2. [45 MARKS]

In this question, you will be implementing the necessary details for a VAE. We will use these implementations to reconstruct the MNIST data set.

1. (10 marks) Complete the `reconstruct` function.

2. (20 marks) Complete the `elbo` loss (10 marks), including the KL-divergence (5 marks) and cross entropy (5 marks) used to compute it.

3. (15 marks) Complete the `train_mnist_vae` function.

Question 3. [40 MARKS]

In this question, you will be looking at what changes occur in a VAE when we condition on the class of the input. First, we will derive the conditional ELBO (CELBO) objective, and then you will implement the necessary details in the code.

(a) [12 MARKS] In a conditional VAE (CVAE), we condition the distributions we want to learn on the label of the data's class. Show that

$$-\ln p(\mathbf{x}|y,\mathbf{W})+D_{KL}(q(\cdot|\mathbf{x},y)||p(\cdot|\mathbf{x},y,\mathbf{W})) = D_{KL}(q(\cdot|\mathbf{x},y)||p)-\mathbb{E}_{\mathbf{h}\sim q(\cdot|\mathbf{x},y)}[\ln(p(\mathbf{x}|\mathbf{h},y,\mathbf{W}))] \quad (2)$$

where $\mathbf{x}$ are the inputs, y is the class label of $\mathbf{x}$, $\mathbf{W}$ are the weights, $q(\mathbf{h}|\mathbf{x},y)$ is the *variational* distribution (learned by the encoder), and $p(\mathbf{x}|\mathbf{h},y,\mathbf{W})$ is the data distribution (learned by the decoder).

(b) [3 MARKS] Given the formula in Equation (2), what is the conditional evidence lower bound objective (CELBO)?

(c) [25 MARKS] Complete the implementation for the **CVAE**.

1. (8 marks) Complete the `reconstruct` function
2. (9 marks) Complete the `celbo` loss
3. (8 marks) Complete the `train_mnist_cvae` function

(d) [5 MARKS] In the final section, we present several visualizations. After showing the reconstruction visualization, we use the two models to generate data without relying on new data for reconstruction. Use the CVAE visualization and change the class that we are generating. In your opinion, which MNIST digits are the hardest to generate and why? Please save some plots and use them to justify your answer.

Note that many answers are acceptable; there is no single digit that is universally the hardest to generate. The primary goal of this question is to have you examine the sampled images and report your observations.

Homework policies:

Your assignment should be submitted on eClass as a single pdf document and a zip file containing: the code and a pdf of written answers. The answers must be written legibly and scanned or must be typed (e.g., Latex).

We will not accept late assignments. Plan for this and aim to submit at least a day early. If you know you will have a problem submitting by the deadline, due to a personal issue that arises, please contact the instructor as early as possible to make a plan. If you have an emergency that prevents submission near the deadline, please contact the instructor right away. Retroactive reasons for delays are much harder to deal with in a fair way.

All assignments are individual. All the sources used for the problem solution must be acknowledged, e.g. web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously; for detailed information see the University of Alberta Code of Student Behaviour.

Good luck!