HY673 - Assignment #2

Yannis Pantazis

Monday March 20th, 2023
Due date: Monday April 3rd, 2023

Instructions

- **Due date:** Monday April 3rd, 2023
- Submission via e-mail to the class account: hy673@csd.uoc.gr
- Provide one file with the solutions.
- Provide one folder with code.
  - The name of each file in the folder should indicate the respective exercise (e.g., ex_2c.py or exercise_2c.py).

Exercise 1: Implementation of RealNVP on MNIST dataset

In this exercise, you will extend the NICE model which was presented in Tutorial 4.
(a) Implement RealNVP model. Essentially, you have to implement

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} = \begin{bmatrix} z_A \\ z_B \odot e^{\alpha_\theta(z_A)} + m_\theta(z_A) \end{bmatrix}$$

where $\{x_A, x_B\}$, $\{z_A, z_B\}$ are partitions of the data and noise vectors, respectively, $\odot$ denotes element-wise product and both $m_\theta(\cdot)$ and $\alpha_\theta(\cdot)$ are neural networks with parameters θ. Note that $\dim(x_A) = \dim(z_A)$ and $\dim(x_B) = \dim(z_B)$.

(b) Train RealNVP on MNIST dataset. Keep in mind that the above RealNVP equation applies to all transformation steps. Experiment with 5 and 10 coupling layers and compare the results.

(c) Perform linear and sinusoidal interpolations between two MNIST digits in the latent space (i.e., the base space of z). In particular, let $z^{(1)}, z^{(2)}$ be two

MNIST digits in the latent space, you will generate and plot
$z_\lambda = (1-\lambda)z^{(1)} + \lambda z^{(2)}$ (linear) and
$\bar{z}_\lambda = (1-\sin(\lambda\pi/2))z^{(1)} + \sin(\lambda\pi/2)z^{(2)}$ (sinusoidal)
interpolations for $\lambda = 0 : 0.1 : 1$.

Exercise 2: Conditional RealNVP on MNIST

(a) Add digit information to the RealNVP model and learn the conditional distributions. In particular, use one-hot encoding for the digit labels and concatenate them in the input of the neural nets. The equation of the conditional RealNVP model is given by

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} = \begin{bmatrix} z_A \\ z_B \odot e^{\alpha_\theta(z_A, y)} + m_\theta(z_A, y) \end{bmatrix}$$

where $y \in \mathbb{R}^{10}$ corresponds to the one-hot encoding vector of the MNIST digit labels.

(b) Train conditional RealNVP on MNIST dataset. Compare the obtained results with the generated digits from 1(b). Can we say something about the number of required transformations when conditional generation is utilized?

Exercise 3: Coupling layers with triple split

In this exercise we will generalize NICE in a different way. Assume that we break both x and z into three groups: $x = \begin{bmatrix} x_A \\ x_B \\ x_C \end{bmatrix}$ and $z = \begin{bmatrix} z_A \\ z_B \\ z_C \end{bmatrix}$ with $\dim(x_A) = \dim(z_A)$, $\dim(x_B) = \dim(z_B)$ and $\dim(x_C) = \dim(z_C)$ and define the transformation

$$x = f(z) := \begin{bmatrix} z_A \\ z_B + m(z_A) \\ z_C + l(z_A, z_B) \end{bmatrix} .$$

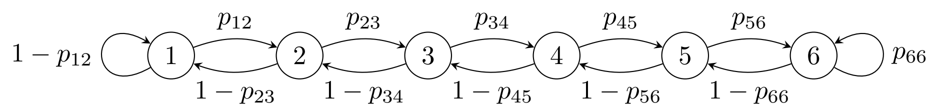
(a) Calculate the inverse transformation.

(b) Calculate the determinant of the Jacobian of f.

(c) Implement this 3-split coupling layer on MNIST data. Provide details regarding masking (how many masks did you use?), permutation of the elements (how was done?), the number of applied transformations. Compare the results with the implementation of NICE from the tutorial.

Exercise 4: Finite-state discrete-time Markov chain

Consider the Markov chain in Figure below which has 6 states.



(a) Write down the transition probability matrix for the Markov chain.

(b) Assume $p_{12} = p_{23} = ... = p_{56} = p_{66} = \frac{1}{2}$ and at time $t = 0$ the initial state is $x_0 = 1$. Simulate (i.e., create a trajectory or a path) the Markov chain using the transition probability matrix for $T = 10^3$ and $T = 10^5$ steps.
Hint: at each step you draw a sample from a categorical distribution whose probabilities correspond to the (transition) matrix row of the current state.

(c) Compute the stationary distribution using eigenvalue decomposition of the transpose of the transition probability matrix and validate the answer using the histogram from the simulation in 4(b).

(d) Re-estimate the stationary distribution when $p_{23} = 0.1$. What do you observe?

(e) Re-estimate the stationary distribution when $p_{12} = p_{23} = ... = p_{56} = p_{66} = 0.1$. What do you observe?

Exercise 5: Compute the Kullback-Leibler divergence rate

Assume a discrete-time stationary Markov process with stationary distribution $\pi(x)$ and true transition probability function $p(y|x)$ where both x and y take values in $\mathcal{X}$ (i.e., $x, y \in \mathcal{X}$). Recall that stationary distribution means that $\pi(y) = \int_{\mathcal{X}} p(y|x)\pi(x)dx$. The stationary (path) distribution for the n-step Markov process $x_{0:n} = (x_0, ..., x_n)$ is $p(x_0, ..., x_n) = \pi(x_0) \prod_{i=1}^{n} p(x_i|x_{i-1})$. Assume also an approximate transition probability function $q(y|x)$ that give raise to the approximate path distribution $q(x_0, ..., x_n) = \pi(x_0) \prod_{i=1}^{n} q(x_i|x_{i-1})$. Compute the limit

$$\lim_{n\to\infty} \frac{1}{n} D_{KL}(p(x_0, ..., x_n)||q(x_0, ..., x_n))$$

Hint: First, write down the definition of Kullback-Leibler divergence for the path distributions, second, exploit the factorization of the path distributions into a product of transition probability functions and finally apply repeatedly the fact that the process is stationary.