NNDL2024: Exercices & Assignments
Session 6 (30 April)

Mathematical Exercises

1. Consider a multivariate Gaussian distribution with covariance $\mathbf{\Sigma}$ and mean $\mathbf{0}$. Recall the score function from the EBM slides.

   (a) (5 points) For algebraic simplicity, show that in this zero-mean Gaussian case, the Langevin iteration can be expressed as

   $$\mathbf{x}_{t+1} = \mathbf{M}\mathbf{x}_t + \alpha\mathbf{n}_t \tag{1}$$

   for some matrix $\mathbf{M}$ and a scalar α. Give those $\mathbf{M}$ and α.

   (b) (5 points) Fix in initial point $\mathbf{x}_0$. Express the results of the *three* first steps $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ of the Langevin algorithm for this Gaussian model, as a function of $\mathbf{x}_0$, the noise terms, $\mathbf{M}$ and α.

   (c) (5 points) From the above, intuitively conclude the formula for $\mathbf{x}_T$ after T steps. (It is enough to intuitively half-guess the formula.)

   (d) (5 points) Consider $\mathbf{\Sigma} = \mathbf{I}$. Show that in the limit T $\to \infty$, the influence of the initial point reduces to zero.

2. Consider an extremely simple latent variable model for which we can do calculations that are usually intractable. The latent variable $z \sim \mathcal{N}(0, 1)$ is a scalar. There is also additive noise $n \sim \mathcal{N}(0, \sigma^2)$. The data is generated as

   $$x = az + n \tag{2}$$

   where a is a scalar parameter. Assume we are given a sample $\mathbf{X} = (x_1, \ldots, x_N)$ of independent observations. To this corresponds an unobserved set $\mathbf{Z} = (z_1, \ldots, z_N)$.

   (a) (5 points) Write explicitly the pdf's $p(z)$ and $p(x|z)$. Give the joint pdf $p(x, z)$.

   (b) (5 points) Here, exceptionally, we can integrate out the latent variable analytically. Thus, derive the marginal probability of x as

   $$p(x) = \int p(x, z)dz \tag{3}$$

Hint: there is no need to explicitly compute the integral; it is enough to use some special properties of the Gaussian distribution, in particular what we know about the distribution of a sum of Gaussians.

(c) (5 points) Based on the above, propose a maximum likelihood estimator for a.

3. (5 points) Show that in a minimax problem, the min and max operators can be interchanged as:

$$\min_a \max_b J(a,b) = -\max_a \min_b -J(a,b) \tag{4}$$

4. (10 points) Assume $\mathbf{x}$ is zero-mean, white and Gaussian. Consider $\mathbf{U}$ an orthogonal matrix. What is the distribution of $\mathbf{y} = \mathbf{U}\mathbf{x}$? (This is proving in detail a results alluded to in the slides.) Hint: there are two ways of approaching this. One is by deriving the pdf of $\mathbf{y}$ using the probability transformation formula, if you happen to know it. The second is, again, using the special properties of the Gaussian distribution...

Computer Assignments

1. (25 points) Consider the bivariate Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$

$$\mathbf{m} = (1,1), \mathbf{\Sigma} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix} \tag{5}$$

Run the Langevin algorithm to sample from this distribution, always starting at $\mathbf{0}$ for simplicity. Try the following specifications:

- Different numbers of iterations: 1,000 ; 10,000 ; 100,000.
- Different step sizes: 0.001,0.01,0.15.

With these $3 \times 3 = 9$ different scenarios, run the algorithm once for each scenario. Visualize the results by plotting corresponding nine 2D histograms, each histogram being based on all the data points one of the scenarios (the code template contains the plotting code). Discuss the results; what hyperparameter choices were successful and why? **Report** the nine histograms and the discussion.

2. (25 points) Load the usual MNIST dataset. Train a GAN on that data. In this final exercise we go towards real life scenarios where you don't know what is a good architecture, so you need to decide the architecture and other details by yourself. Don't try anything too big and ambitious! Also, remember that the lecture slides pointed out that GAN training is quite difficult: if you cannot make it work within a reasonable amount of time, don't waste too much time on it; an honest and correct attempt will give you almost the maximum points. You are allowed to use code downloaded from the Internet if you prefer, and you can also use another

variant of GAN apart from the original one. **Report** ten generated data points. Discuss whether the generation appears successful or not, and if not, what might be the reason.