NNDL2024: Exercices & Assignments
Session 5 (25 Apr)

Mathematical Exercices

1. (10 points) The unit sphere is a simple example of a manifold. Its definition as the set of zeros of a function was given in the slides. Consider the 2D plane ($m = 2$ in the notation of the slides). Give the definition of the unit sphere as the image of a set U going through a function $\mathbf{f}$. What is $\mathbf{f}$, what is U? What is the dimension of the manifold n, that is, of the space where U lies ? (Hint: think about basic trigonometrics.)

2. (5 points) Another well-known manifold is the manifold of orthogonal matrices. Find a function $\mathbf{g}$ defining this manifold as the set where $\mathbf{g}$ is zero.

3. (15 points) Consider the general multivariate Gaussian model with zero mean. Let us derive the score matching objective given as an example in the slides. That is, derive the score matching objective function for estimating the inverse covariance (precision) parameter matrix, fixing the mean to be zero. (Hint: you can use the results at: `https://math.stackexchange.com/q/222894`)

4. Consider a simple model for $x \in \mathbb{R}$ (properly called an exponential family with one sufficient statistics), defined as:

$$\log p_{\text{un}}(x; \theta) = f(x)g(\theta) \tag{1}$$

   for some smooth functions $f, g : \mathbb{R} \to \mathbb{R}$, and $\theta \in \mathbb{R}$.

   (a) (5 points) Derive the score function

   (b) (10 points) Derive the score matching objective

   (c) (5 points) It turns out the maximization of the score matching objective with respect to θ is quite easy (which is a general property of exponential families). Find a simple expression for the optimizing $\hat{\theta}$.

5. (5 points) Suppose you train logistic regression so that the two datasets you're trying to discriminate have equal distributions. Assume the number of data points is equal in the two classes, and approaches infinity. What is the optimal regression function in that case?

6. (5 points) Dr. Crack Pot proposes a variant of NCE where the noise distribution is uniform in $\mathbb{R}^n$. He claims that this is the ideal noise distribution since then the regression function will directly converge to the log-pdf of the real data. What is wrong with his proposal?

Computer Assignment (40 points) Get the MNIST data again. Train a dimension-reducing autoencoder which has only two units in the "main" hidden layer that performs the encoding, i.e. the output of $\mathbf{g}$ in the notation of the slides has a dimension of $m = 2$. Try two different architectures for the encoding and decoding networks, with depths of 2 and 3 layers (same depth for encoding and decoding layers). As activation function, use ReLU, and set the number of units in each of the layers inside the encoding and decoding networks to 10. There are other details to be fixed, but they are left to you. **Report** the resulting two 2D visualizations of the data, so that you plot 500 randomly selected digits on the 2D plane. See the template for the relevant commands.