Deep Learning & Applied AI
2^{nd} Semester 2020/2021

Computer Science
Sapienza University of Rome

Midterm

Thu 08 Apr 2021, 17:00-19:00

Name: ______________________________

Matr. number: ______________

There is a total of **12** questions.

The maximum possible score is **40**.

NB: All answers (including multiple choices) must be motivated.

Any non-motivated answer, even if correct, will be considered wrong.

Question 1 (2 points).

Explain the differences between the transpose and reshape operations, and provide an example on a tensor T.

Question 2 (2 points).

Let us be given two MLPs with k hidden layers, each with dimension d, employing the following activation functions:

1. MLP_1: $f(x) = |x|$
2. MLP_2: $f(x) = 3x + 5$

Do the two networks carry the same expressive power?

Question 3 (2 points).

How many learnable parameters has a 2-layer MLP (input-hidden-output) with 10-42-3 neurons on each layer? What is the size of the class of hypotheses we are considering when we opt for such an architecture? Assume we implement weights as float32.

Question 4 (2 points).

What does the energy landscape of a network depend from?

- ☐ The data
- ☐ Number of layers
- ☐ Loss function
- ☐ Network parameters
- ☐ Activation functions
- ☐ All the above

Question 5 (2 points).

Let us be given a MLP with $k > 1$ hidden layers, each with fixed dimension $\leq d$, with nonlinear activation function σ. Can it learn *any* function?

Question 6 (3 points).

Let us be given a dataset containing all the characters of all existing fonts, in the form of one single 50×50 training image, rasterized using a given function f. On this dataset, we want to train a model that must be able to classify the font used in some test 50×50 image I; we know that the font in I has been rasterized using f. Describe the risk of overfitting in this scenario.

Question 7 (3 points).

Consider the linear regression problem with Tikhonov (L_2) regularization. Find the closed form solution of this regression problem using matrix notation.

Question 8 (3 points).

Let $\mathcal{S}_n = \{N_\Theta\,,\ |\Theta| = n\}$ be the space of "spherical deep nets", defined as the set of deep neural networks, parametrized by Θ, such that the set of weights of each network identifies a point on the n-dimensional unit sphere. How would you train a network M_Θ to ensure that $M_\Theta \in \mathcal{S}_n$?

Question 9 (4 points)**.**

How would you modify SGD with momentum to ensure that the descent speed does not depend on the steepness of the loss function? In this case, would momentum still provide an acceleration?

Question 10 (4 points)**.**

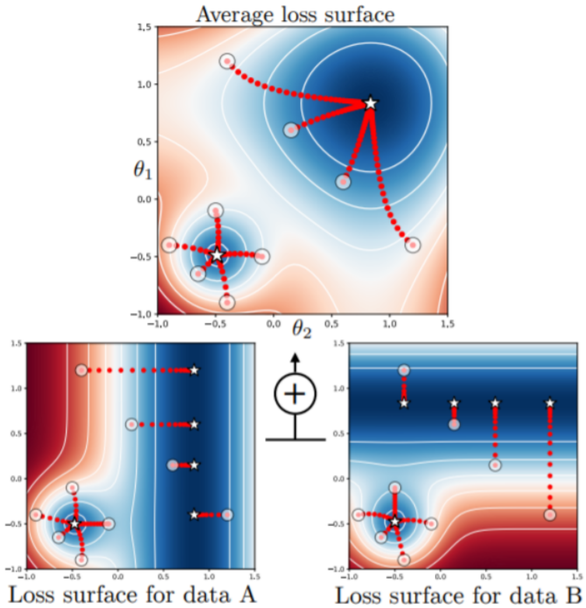
Imagine you are solving a digit classification problem using a MLP with 5 layers. Consider the following three scenarios:

1. Your dataset is the vanilla MNIST.
2. Your dataset is MNIST, but each 28×28 image is multiplied by the same 28×28 matrix $\mathbf{T}$, where $\mathbf{T1} = \mathbf{1}$, $\mathbf{1}^\top\mathbf{T} = \mathbf{1}^\top$, and $\mathbf{T}^\top\mathbf{T} = \mathbf{I}$.
3. Your dataset is MNIST, but each 28×28 image is multiplied by the same 28×28 matrix $\mathbf{T}'$, where $\mathbf{T}'$ is composed by the first 14 columns of $\mathbf{T}$ repeated twice.

In which scenario do you expect your model to perform the best? In which one do you expect it to perform the worst?

Question 11 (5 points).

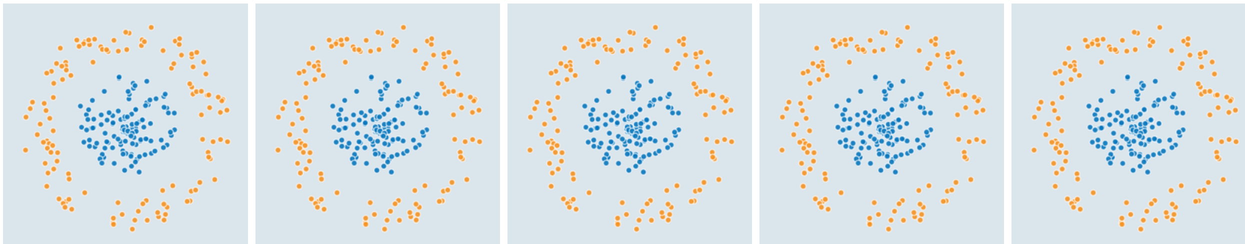
Consider the following illustration:



Explain, to the best of your understanding, what are the key messages that the image above is trying to convey. Elaborate as much as you can.

Question 12 (8 points).

We are given a dataset consisting of 400 2D points for a binary classification task (0: orange, 1: blue), which we plot below five times. Use each image to answer the five questions below.



(1) (2) (3) (4) (5)

1. On image (1), draw the best decision boundary achievable by a MLP with 1 layer: 2-1 (input-output) and trained taking as input the point coordinates (x_1, x_2).

2. On image (2), draw the best decision boundary achievable by a MLP with 2 layers: 2-2-1 and ReLU activations, trained taking as input the point coordinates (x_1, x_2).

3. On image (3), draw the best decision boundary achievable by a MLP with 3 layers: 2-3-1 and ReLU activations, trained taking as input the point coordinates (x_1, x_2).

4. On image (4), draw the best decision boundary achievable by a MLP with 3 layers: 2-3-1 and **linear** activations, trained taking as input the point coordinates (x_1, x_2).

5. How can you achieve a better result for the latter network without changing architecture (same number of neurons and same activation functions)? On image (5), draw the best decision boundary achievable with your approach.