Columbia University
IEOR4742 – Deep Learning for OR & FE (Hirsa)
Assignment 1 – Due 11:40 am on Tuesday February 25th, 2020

Problem 1 (Impact of non-linear activation functions on Learning): Show that a feed-forward neural network with linear activation function and any number of hidden layers is equivalent to just a linear neural neural network with no hidden layer.

Problem 2 (Linear Classification Example): Assume $\Omega = [-100, 100] \times [-100, 100]$. Define the following two spirals in the two-dimensional space Ω

$$\begin{aligned} x_1 &= r_1 \cos(\phi_1) \\ y_1 &= r_1 \sin(\phi_1) \end{aligned}$$

$$\begin{aligned} x_2 &= r_2 \cos(\phi_2) \\ y_2 &= r_2 \sin(\phi_2) \end{aligned}$$

where

$$\begin{aligned} r_1 &= 60 + 0.20t \\ r_2 &= 40 + 0.40t \\ \phi_1 &= -0.06t + 3 \\ \phi_2 &= -0.08t + 3 \end{aligned}$$

and $t = 1, 2, \ldots, 100$

(a) plot the two spirals

(b) for classification, transform the 2-dim space to a 3-dim space and find a hyperplane to classify, use hinge loss to achieve this

Problem 3 (Linear Classification Example): Assume $\Omega = [-1, 1] \times [-1, 1]$. Define the following two curves in the two-dimensional space Ω

$$\begin{aligned} y_1 &= -0.6 \sin(\pi/2 + 3x) - 0.35 \\ y_2 &= -0.6 \sin(\pi/2 + 3x) + 0.25 \end{aligned}$$

(a) In the first trial, simply separate two classes of data (two curves) by dividing them with a straight line.

(b) Now transforming the space Ω using the following transformation to create new presentation for the space, that is for every grid point in Ω we do

$$\begin{aligned} \hat{x} &= \tanh(w_{11}x + w_{21}y + b1) \\ \hat{y} &= \tanh(w_{12}x + w_{22}y + b2) \end{aligned}$$

for some parameter set $(w_{11}, w_{12}, w_{21}, w_{22}, b_1, b_2)$. One can run the following nested loop for the parameter set to visualize the transformed topology:

```
for w11 = -3, ..., 3 do
    for w12 = -3, ..., 3 do
        for w21 = -3, ..., 3 do
            for w22 = -3, ..., 3 do
                for b1 = -1, ..., 1 do
                    for b2 = -1, ..., 1 do
                        x̂ = tanh(w11 x + w21 y + b1)
                        ŷ = tanh(w12 x + w22 y + b2)
                    end
                end
            end
        end
    end
end
```

It is clear there are infinitely many presentations. We are looking for the one such that one can just draw a line through the transformed data without crossing any of the `transformed` curves. Note that we seek $\mathbf{w} = (a, b, c)$ as well as $(w_{11}, w_{12}, w_{21}, w_{22}, b_1, b_2)$ (9 parameters) such that

$$\mathbf{w}^\top \eta \geq 0 \text{ when } t = +1$$
$$\mathbf{w}^\top \eta < 0 \text{ when } t = -1$$

or equivalently

$$\mathbf{w}^\top \eta_j t_j \geq 0 \quad \forall j$$

thus we seek to minimize

$$-\sum_{j \in \mathcal{A}} \mathbf{w}^\top \eta_j t_j$$

where $\mathcal{A}$ is the set of mis-classified inputs. Here $\eta_j = (\hat{x}_j, \hat{y}_j, 1)$.

(c) Use hinge loss to find the optimal parameter set

in all cases examine sensitivity of the obtained optimal parameter set with respect to the starting point