IEC-03 AI Techniques (2024)
Assignment -2 (Deadline for submission is Oct 18 (in tutorial & in person) (no further extension is allowed))
Instructions: All solutions are to be submitted in hardcopy, while the solutions to programming questions must be demonstrated in person during the Friday tutorial

Q.1 Explain the Elbow Method in the context of unsupervised learning. How is it used to determine the optimal number of clusters in a dataset when applying k-means clustering?

Q.2 Consider the case of the XOR function in which the two points $\{(0, 0),(1, 1)\}$ belong to one class, and the other two points $\{(1, 0),(0, 1)\}$ belong to the other class. Show how you can use the ReLU activation function to separate the two classes.

Q.3 Show the following properties of the sigmoid and tanh activation functions (denoted by $\Phi(\cdot)$ in each case):

(a) Sigmoid activation: $\Phi(-v)=1 - \Phi(v)$

(b) Tanh activation: $\Phi(-v) = -\Phi(v)$

(c) Hard tanh activation: $\Phi(-v) = -\Phi(v)$ (where $\Phi(v) = \max \{\min [v, 1] , -1\}$ (hard tanh))

Q.4 Show that the tanh function is a re-scaled sigmoid function with both horizontal and vertical stretching, as well as vertical translation:

$\tanh(v) = 2*\text{sigmoid}(2v) - 1$

Q.5 Consider a network with two inputs x1 and x2. It has two hidden layers, each of which contain two units. Assume that the weights in each layer are set so that the top unit in each layer applies sigmoid activation to the sum of its inputs and the bottom unit in each layer applies tanh activation to the sum of its inputs. Finally, the single output node applies ReLU activation to the sum of its two inputs. Write the output of this neural network in closed form as a function of x1 and x2. This question should give you an idea of the complexity of functions computed by neural networks.

Q.6 Compute the partial derivative of the closed form computed in the previous exercise with respect to x1. Is it practical to compute derivatives for gradient descent in neural networks by using closed-form expressions (as in traditional machine learning)?

Q. 7 Consider the following recurrence: $(x(t+1), y(t+1))=(f(x(t), y(t)), g(x(t), y(t)))$

Here, f() and g() are multivariate functions. (a) Derive an expression for $\partial x(t+2)/ \partial x(t)$ in terms of only x(t) and y(t). (b) Can you draw an architecture of a neural network corresponding to the above recursion for t varying from 1 to 5? Assume that the neurons can compute any function you want.

Q.8 For a one-dimensional time series of length L and a filter of size F, what is the length of the output?

Programming questions (demonstrate the code to TA in person, and submit code, solution to questions separately)

Q.8 Implement and demonstrate (in person)

(a) Python code for backpropagation for a Multi-layered Neural Network (MLN)?

(b) Use the above developed module to implement the Python code on MLN for XOR, NAND, OR, AND gates. Plot error function as a function of number of iterations for all the cases above and save the files.

Q.9 Consider a data set in which the two points $\{(-1, -1),(1, 1)\}$ belong to one class, and the other two points $\{(1, -1),(-1, 1)\}$ belong to the other class. Start with perceptron parameter values at (0, 0), and work out a few updates with α (also called eta) = 1. While performing the learning updates, cycle through the training points in any order.

(a) Does the algorithm converge in the sense that the change in objective function becomes extremely small over time? (b) Explain why the situation in (a) occurs.

$\Phi(v) = \max\{v, 0\}$ (Rectified Linear Unit [ReLU])

$\Phi(v) = \max \{\min [v, 1] , -1\}$ (hard tanh)

Q.10 Download an implementation of the AlexNet architecture from a neural network library of your choice. Train the network on subsets of varying size from the ImageNet data, and plot the top-5 error with data size.