ELL409: Assignment 2

Demo Schedule: 1st week of May 2021

Neural Networks

In this assignment you will implement a neural network to recognize handwritten digits using the same data set that was used in Assignment 1 (`https://web.iitd.ac.in/~seshan/a1/handwritten_image_data.rar`). The neural network will be able to represent more complex models that form non-linear hypotheses. You are expected to implement the forward propagation algorithm to determine the activations of the nodes in the neural network and the backpropagation algorithm for learning the neural network parameters. Do not use any of the built-in neural network model libraries.

- You can assume a neural network with only 3 layers.
- Implement the feedforward computation that computes $h(x^{(i)})$ for every training example i and sum the cost over all examples.
- When training neural networks, the parameters are randomly initialised. One strategy is to randomly select values for $\mathbf{W}^{(l)}$ uniformly in the range $[-\epsilon, \epsilon]$ where $\epsilon = \frac{\sqrt{6}}{\sqrt{S_l + S_{l+1}}}$ and S_l, S_{l+1} are the number of neurons in the layers adjacent to $\mathbf{W}^{(l)}$.
- Implement the backpropagation algorithm by computing the error terms associated with the output layer and propagating these errors back towards the input layer. Use the following expression for the hidden layers:

$$\boldsymbol{\delta}^{(l)} = (\mathbf{W}^{(l)})^T \boldsymbol{\delta}^{(l+1)} .* g\prime(\mathbf{z}^{(l)})$$

  where $.*$ represents elementwise multiplication and $g\prime(z) = g(z)(1 - g(z))$ is the derivative of the sigmoid function.
- The gradient of the neural network cost function is given by:

$$\frac{\partial E(\mathbf{W})}{\partial \mathbf{W}_{ij}^{(l)}} = \frac{1}{N} \boldsymbol{\Delta}_{ij}^{(l)}$$

  where $\boldsymbol{\Delta}^{(l)} := \boldsymbol{\Delta}^{(l)} + \boldsymbol{\delta}^{(l+1)}(\mathbf{a}^{(l)})^T$.
- Learn a good set of parameters for the neural network by providing the neural network cost function and gradient to an unconstrained optimisation routine. Apply the learned model to determine the training set accuracy.
- Now introduce regularisation and comment on the differences. Note that with regularisation an additional term is introduced in the gradient computation:

$$\frac{\partial E(\mathbf{W})}{\partial \mathbf{W}_{ij}^{(l)}} = \frac{1}{N} \Delta_{ij}^{(l)} + \frac{\lambda}{N} \mathbf{W}_{ij}^{(l)} \text{ for } j \geq 1.$$

You should prepare a report, compiling all your results and your interpretation of them, along with your overall conclusions. In particular, you should attempt to answer all of the questions posed above. Any graphs or other visualisations should also be included in the report.