Assignment 2: Backpropagation / Neural Network

Deep Networks for Machine Learning — Spring 2019

Brief

• Due date: 23 March, 2019 11:59 PM
• Implement linear regression
• Required files: Submit your ipython notebook along with a Report.pdf in a single zip file with name IDnumber_A02.zip.
• Submission: iLearn

Overview

The goal of this assignment is to derive the equations for backpropagation algorithm for a three layer fully connected neural network. We will then also use keras to implement a multilayer neural network report performance

Task 1 (15 points)

The goal of this part to develop a theoretical understanding of how do you get the expressions with backpropagation algorithm. Suppose you have a three layer fully connected neural network (input layer, hidden layer and output layer). Following are some further "specifications":

- The input feature vectors are d dimensional (you can think of these as MNIST images, flattened as vectors)
- There are N feature vectors in you training dataset
- There are H hidden nodes and K output nodes (for K mutually exclusive classes).
- This neural network is to be trained for classification under cross-entropy loss function

Derive the equations for batch gradient updates for the input-to-hidden unit weights and hidden-to-output unit weights. Is it wise to use a learning rate parameter? Why?

Task 2 (15 points)

You will be using the fashion mnist dataset available in keras. The dataset can be imported using the following code instance:

```
from keras.datasets import fashion_mnist

(train_images, train_labels), (test_images, test_labels) = fashion_mnist.load_data()
```

Divide the training images further, into 40,000 training and 20,000 validation images. Make sure you do this at random. So now you will have 40,000 training images, 20,000 validation images and 10,000 testing images.

1. Read about the fashion mnist dataset and report the highlights (e.g. number of images, number of classes etc) along with some example images.

2. What will be reasonable loss function to use for classification if the last layer activation functions are **soft-max** functions? Explain your rationale.
3. Train four neural networks each with one hidden layer; one with 16 nodes, one with 64 nodes, one with 128 node and one with 512 nodes. Train each one of them for 15 epochs.

    - Which one of them gives you the best performance on the validation data?
    - Does increasing epoch to 30 improve the validation performance?
    - Experiment with using sigmoid and relu as activation functions for the hidden layer. Plot the validation loss or validation accuracy for both kinds of activation functions for all the four network models against the number of epochs (30). This will give you 8 plots. Comment on the plots and comparison across the choice of activation functions
    - Pick the model parameters that give you the best performance, then combine your validation data with training data, retrain your network with optimum parameters and report the classification accuracy as well as the confusion matrix on the test data.

Task 3 (15 points)

Answer briefly in one or two paragraphs.

1. What is generalization in the context of machine learning. Is it better for the machine learning algortihms to generalize?
2. What is overfitting and underfitting, how is it related to generalization?
3. What is capacity of a neural network? Is it related to generalization, if so how?
4. What is the relationship between number of epochs and overfitting?
5. What is weight regularization? How does it improve generalizations? Write a code snippet that can be included in Task 2 for adding weight regularization.
6. What is drop out? How does it improve generalizations? Write a code snippet that can be included in Task 2 for adding drop out.

Task 4 (Reading assignment)

1. http://cs231n.github.io/optimization-1/
2. http://cs231n.github.io/neural-networks-1/
3. http://cs231n.github.io/neural-networks-2/
4. http://cs231n.github.io/neural-networks-3/

Honor code

• Honor code of AUS will be followed.
• Always give reference for code or ideas.
• Do not copy code from the web or from any other person.

Submission

• Please note that the zip file must be named in following manner IDnumber_A02.zip (e.g. g00012345_A01.zip).
• No late submissions will be accepted.