ECE 520.438 & 520.638: Deep Learning
Homework 2: Perceptron, Logistic Regression
Spring 2024

1 Part 1: (20 points)

Answer the following questions.

1. What kind of decision boundary is learned by perceptrons?
2. How can we use perceptrons to do C-class calssfication?
3. Breifly explain stochastic gradient descent.
4. Briefly explain mini-batch gradient descent.
5. Let $\sigma(a) = \frac{1}{1+e^{-a}}$. Show that $\sigma'(a) = \sigma(a)[1 - \sigma(a)]$.
6. What is an Epoch?
7. If the dataset size is 2000 samples and the batch size is 50, then an epoch will consists of how many iterations?
8. Breifly describe Dropout.
9. What is early stopping?
10. Why is regularization used during training a neural network?
11. What are different ways in which we can reduce over-fitting in neural network?
12. Why do we use batch normalization?

2 Part 2: Perceptron (40 points)

In this problem, you will classify handwritten digits using the MNIST Database.
http://yann.lecun.com/exdb/mnist/
Please download the four files found there, these will be used for this homework. To reduce computation time, please use only the first 20,000 training images/labels and only the first 2,000 testing images/labels.

Read in the data from files. Each image is 28×28 pixels. An image can be viewed as a 784-dimensional vector of pixel intensities. For each image, append a '1' to the beginning of each x-vector; this will act as an intercept term (for bias). Use the gradient descent algorithm for Perceptron derived in class, to classify given $x \in \mathbb{R}^{785}$ whether a digit's label is k or if it is some "other" digit, i.e. $k \notin \{0, \cdots, 9\}$. For instance, if you are classifying "2", you would designate $y = 2$ as the positive class, and all other digits as the negative class. For each image, do this two-way classification for all 10 digits. You will train 10 perceptrons that will collectively learn to classify the handwritten digits in the MNIST dataset. Each perceptron will have 785 inputs and one output.

1. Report the test accuracy for each of the 10 2-way classifications on the test set.

2. For each image in the test set, report the overall test accuracy. An example will be considered labeled correctly if the perceptron classification of the true label has the highest probability. So for instance, if the true label was {2} for an image, you would

count it as correctly classified if the perceptron test of $\{2\}$ vs. $\{0, 1, 3, 4, 5, 6, 7, 8, 9\}$ had the highest probability of all the 10 2-way classifications.

3   Part 3: Logistic Regression (40 points)

Logistic regression is a binary classification method which can be modeled as using a single neuron reading in an input vector $x \in \mathbb{R}^d$ and parameterized by weight vector $w \in \mathbb{R}^d$, where the neuron outputs the probability of the class being $y = 1$ given x

$$P(y = 1|x) = g_w(x) = \frac{1}{1 + \exp(-w^T x)} = \sigma(w^T x)$$

$$P(y = 0|x) = 1 - P(y = 1|x) = 1 - g_w(x).$$

Given $\{(x^{(i)}, y^{(i)})\}_{i=1}^N$, the Cross Entropy Loss function is defined as follows

$$J(w) = -\sum_{i=1}^N \left(y^{(i)} \log(g_w(x^{(i)})) + (1 - y^{(i)}) \log(1 - g_w(x^{(i)})) \right),$$

where N denotes the total number of training samples. We will optimize this cost function via gradient descent.

1. Show that the gradient of the cost function with respect to the parameter w is:

$$\frac{\partial J(w)}{\partial w_j} = \sum_{i=1}^N x_j^{(i)} (g_w(x^{(i)}) - y^{(i)}).$$

   Show your work.

2. Using the gradient derived for Logistic Regression cross entropy loss, use gradient descent to classify given $x \in \mathbb{R}^{785}$ whether a digit's label is k or if it is some "other" digit, i.e. $k \notin \{0, \cdots, 9\}$. Report the test accuracy for each of the 10 2-way classifications on the test set.

3. For each image in the test set, report the overall test accuracy. An example will be considered labeled correctly if the perceptron classification of the true label has the highest probability. So for instance, if the true label was $\{2\}$ for an image, you would count it as correctly classified if the perceptron test of $\{2\}$ vs. $\{0, 1, 3, 4, 5, 6, 7, 8, 9\}$ had the highest probability of all the 10 2-way classifications.

Grading: You will be graded based on the code you develop, plus your homework report summarizing your findings. If possible, please write your report using LaTeX.