CSE4088 Introduction to Machine Learning

Homework 3 Report

- **Gradient Descent**

Consider the nonlinear error surface $E(u,v) = (ue^v - 2ve^{-u})^2$. We start at the point $(u,v) = (1,1)$ and minimize this error using gradient descent in the uv space. Use $\eta = 0.1$ (learning rate, not step size).

4. The partial derivative of $E(u,v)$

$$\frac{\partial E}{\partial u} = 2(ue^v - 2ve^{-u})(ue^v - 2ve^{-u})'$$

$$\frac{\partial E}{\partial u} = 2(ue^v - 2ve^{-u})(e^v + 2ve^{-u})$$

The answer is [e].

5. To find minimum value of $E(u,v)$, *scipy.optimize.fmin* library is used. Applying the Gradient Descent Algorithm, expected to find the minimum value.

$$\min\left(E(u,v)\right) = 3.5113320785840053\text{e} - 16$$

$$\min(u) = 0.46163107$$

$$\min(v) = 1.09795193$$

Program output:

```
The minimum value of E(u,v):  3.5113320785840053e-16
u: 0.46163106570457335, v: 1.0979519282628663

Achieved to expected error rate...
The achieved minimum value of E(u,v):  1.2086833944220747e-15
u: 0.04473629039778207, v: 0.023958714099141746
Iteration: 10
```

The iteration number is 10. The answer is [d].

6. According to output of the Problem 5, the closest point to the final (u, v), $(0.045, 0.024)$.

The answer is [e].

7. The algorithm is a little bit changed. Iteration number restricted with 15.

Program output:

```
u: 6.29707589930517, v: -2.852306954077811
Iteration: 15
Current error rate: 0.13981379199615288
```

The error rate is closed to 10^{-1}. The answer is [a].

- **Logistic Regression**

In this problem you will create your own target function f (probability in this case) and data set D to see how Logistic Regression works. For simplicity, we will take f to be a 0/1 probability so y is a deterministic function of x.

8. In this question, we applied Stochastic Gradient Descent and estimated the E_{out}. Repeated the experiment 100 times and found the average E_{out}.

Program output:

```
Sample 1, Epoch: 384, E_out: 0.1807
Sample 2, Epoch: 419, E_out: 0.1719
Sample 3, Epoch: 409, E_out: 0.1691
...
Sample 98, Epoch: 211, E_out: 0.1153
Sample 99, Epoch: 335, E_out: 0.1408
Sample 100, Epoch: 398, E_out: 0.1486
Average E_out:  0.09428157166985335
Average Epochs:  337.64
```

The average $E_{out} = 0.09428$. The answer is [d].

9. According to results in question 8, we found also average epoch.

The average epochs = 337.64. The answer is [a].

- **Regularization with Weight Decay**

In this problem, we will use training and test data from file. We are going to apply Linear Regression with a non-linear transformation for classification. The nonlinear transformation is given by

$$\Phi(x_1, x_2) = (1,\ x_1,\ x_2,\ x_1^2,\ x_2^2, x_1x_2, |x_1 - x_2|, |x_1 + x_2|)$$

Recall that the classification error is defined as the fraction of misclassified points.

2. Run Linear Regression on the training set after performing the non-linear transformation. What values are closest to the in-sample and out-of-sample classification errors, respectively?

Program output:

```
Linear Regression => training error: 0.028571, test error: 0.084000
```

The closest values are 0.03, 0.08. The answer is [a].

3. Add weight decay to Linear Regression. $\lambda = 10^k$. For $k = -3$

Program output:

```
Linear Regression, k=-3 =>  training error: 0.028571, test error: 0.080000
```

The closest values are 0.03, 0.08. The answer is [d].

4. $\lambda = 10^k$. For $k = 3$

Program output:

```
Linear Regression, k=3 =>  training error: 0.428571, test error: 0.472000
```

The closest values are 0.4, 0.4. The answer is [e].

5. What value of k in [-2, 2] achieves the smallest out-of-sample classification error?

Program output:

```
Linear Regression, k=-2 =>  training error: 0.028571, test error: 0.080000
Linear Regression, k=-1 =>  training error: 0.028571, test error: 0.056000
Linear Regression, k=0 =>  training error: 0.000000, test error: 0.088000
Linear Regression, k=1 =>  training error: 0.057143, test error: 0.116000
Linear Regression, k=2 =>  training error: 0.257143, test error: 0.264000
```

The smallest out-of-sample error is 0.056. k is -1. The answer is [d].

6. What value is closest to the minimum out-of-sample classification error achieved by varying k?

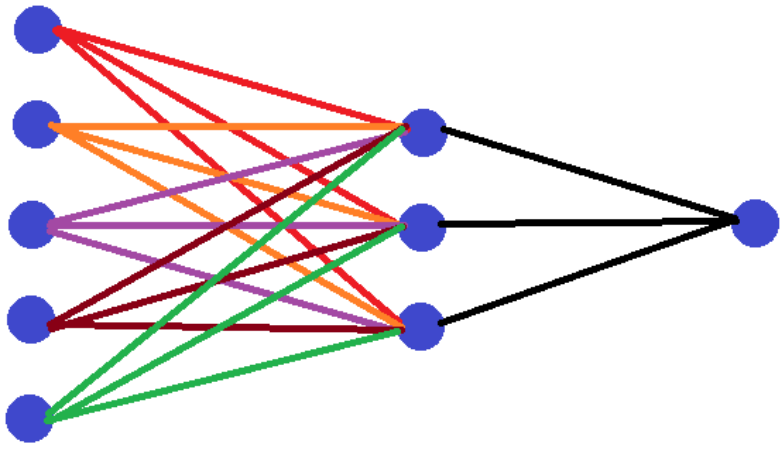
According to results in question 5, the minimum out-of-sample error is 0.056.

The closest answer is 0.6. The answer is [b].

- **Neural Networks**

A fully connected Neural Network has $L = 2,\ d^{(0)} = 5,\ d^{(0)} = 3,\ d^{(2)} = 1$. If only products of the form $w_{ij}^{(l)} x_i^{(l-1)},\ w_{ij}^{(l)} \delta_j^{(l)}, x_i^{(l-1)} \delta_j^{(l)}$ count as operations.

8. What is the total number of operations in a single iteration of backpropagation?



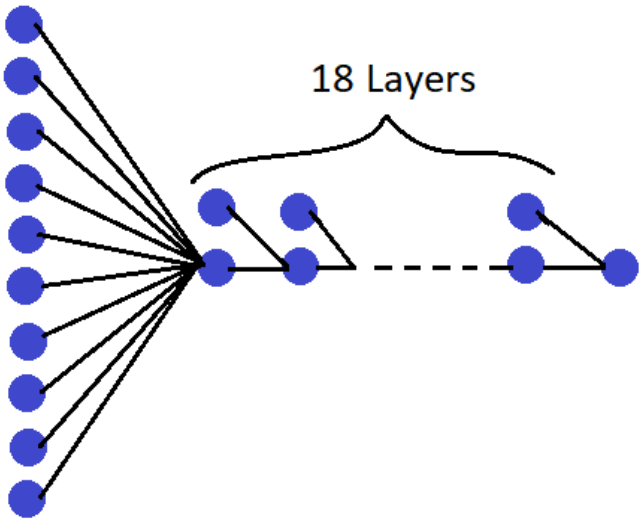
Forward calculations: $w_{ij}^{(l)} x_i^{(l-1)}$ $\quad 5 * 3 + 3 * 1 = 18\ operations$

Backward calculations: $w_{ij}^{(l)} \delta_j^{(l)}$ $\quad 5 * 3 + 3 * 1 = 18\ operations$

Updating weights: $x_i^{(l-1)} \delta_j^{(l)}$ $\quad 5 + 3 + 1 = 9\ operations$

Total 45 operations. The answer is [d].

9. What is the minimum possible number of weights that such a network can have?



When we connect all units sequentially, we find the minimum number of weights.

$$10 * 1 + 18 * 2 = 46\ weights$$

Minimum 46 weights. The answer is [a].

10. What is the maximum possible number of weights that such a network can have?

We have 10 input nodes. To reach maximum network, we put 22 nodes on the first layer, 13 nodes on the second layer.

$$10 * 21 + 22 * 13 + 14 * 1 = 510\ weights$$

The maximum number of weights is 510. The answer is [e].