CS 5350/6350: Machine Learning Spring 2020

Homework 5
Britton Gaul
u0915408

Handed out: 7 Apr, 2020
Due date: 11:59pm, 24 Apr, 2020

1 Paper Problems [40 points]

1.

Derivative of sigmoid function:
$\sigma(s) = \frac{1}{1+e^{-s}} = \frac{\partial \sigma}{\partial s} = \sigma(s)(1-\sigma(s))$

$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y_1} \cdot \frac{\partial y_1}{\partial x} + \frac{\partial z}{\partial y_2} \cdot \frac{\partial y_2}{\partial x} + \frac{\partial z}{\partial y_3} \cdot \frac{\partial y_3}{\partial x}$
$= 3\sigma(y_1)(1-\sigma(y_1)) - e^{-x}\sigma(y_2)(1-\sigma(y_2)) + cos(x)\sigma(y_3)(1-\sigma(y_3))$
$= 3\sigma(3x)(1-\sigma(3x)) - e^{-x}\sigma(e^{-x})(1-\sigma(e^{-x})) + cos(x)\sigma(sin(x))(1-\sigma(sin(x)))$
so,
$\frac{\partial z}{\partial x}_{when,x=0} = 3.3034$

2.

Layer 1:
$z_1^1 = \sigma(\sum_{i=0}^{2} w_{i1}^1 x_i) = \sigma(-1-2-3) = 0.0025$
$z_2^1 = \sigma(\sum_{i=0}^{2} w_{i2}^1 x_i) = \sigma(1+2+3) = 0.9975$
Layer 2:
$z_1^2 = \sigma(\sum_{i=0}^{2} w_{i1}^2 z_i^1) = \sigma(-1-2\cdot 0.0025 - 3\cdot 0.0075) = 0.0180$
$z_2^2 = \sigma(\sum_{i=0}^{2} w_{i2}^2 z_i^1) = \sigma(1+2\cdot 0.0025 + 3\cdot 0.0075) = 0.9820$
Layer 3:
$y = \sum_{i=0}^{2} w_{i1}^3 z_i^2 = -1 + 2\cdot 0.018 - 1.5\cdot 0.982 = -2.4370$

3.

Layer 3 weights:
$\frac{\partial L}{\partial w_{01}^3} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial w_{01}^3} = (y - y*)$
$z_0^2 = -2.4370 - 1 = -3.4370$ this value gets cached
$\frac{\partial L}{\partial w_{11}^3} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial w_{11}^3} = -3.4370 z_1^2 = -3.4370 \cdot 0.018 = -0.0619$
$\frac{\partial L}{\partial w_{21}^3} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial w_{21}^3} = -3.4370 z_2^2 = -3.4370 \cdot 0.982 = -3.3751$
Layer 2 weights:
$\frac{\partial L}{\partial w_{01}^2} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z_1^2} \cdot \frac{\partial z_1^2}{\partial w_{01}^2} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z_1^2} \cdot \frac{\partial \sigma(s)}{\partial s} \cdot \frac{\partial s}{\partial w_{01}^2} = (y - y*) w_{11}^3 \sigma(s)(1-\sigma(s)) z_0^1$ this value is

cached
$= (y - y*)w^3_{11} z^2_1 (1 - z^2_1) z^1_0 = -3.4370 \cdot 2 \cdot 0.018 \cdot (1 - 0.018) = -0.1215$ where this value is also cached
$\frac{\partial L}{\partial w^2_{11}} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z^2_1} \cdot \frac{\partial z^2_1}{\partial w^2_{11}} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z^2_1} \cdot \frac{\partial \sigma(s)}{\partial s} \cdot \frac{\partial s}{\partial w^2_{11}} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z^2_1} \cdot \frac{\partial \sigma(s)}{\partial s} z^1_1$
$= -0.1215 \cdot 0.0025 = -3.0375e - 04$
$\frac{\partial L}{\partial w^2_{21}} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z^2_1} \cdot \frac{\partial z^2_1}{\partial w^2_{21}} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z^2_1} \cdot \frac{\partial \sigma(s)}{\partial s} \cdot \frac{\partial s}{\partial w^2_{21}} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z^2_1} \cdot \frac{\partial \sigma(s)}{\partial s} z^1_2$
$= -0.1215 \cdot 0.0075 = -0.1212$
$\frac{\partial L}{\partial w^2_{02}} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z^2_2} \cdot \frac{\partial z^2_2}{\partial w^2_{02}} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z^2_2} \cdot \frac{\partial \sigma(s)}{\partial s} \cdot \frac{\partial s}{\partial w^2_{02}}$
$= (y - y*)w^3_{21} z^2_2 (1 - z^2_2) z^1_0 = -3.4370 \cdot -1.5 \cdot 0.982 \cdot (1 - 0.982) \cdot 1 = 0.0911$ this value is cached
$\frac{\partial L}{\partial w^2_{12}} = (y - y*)w^3_{21} z^2_2 (1 - z^2_2) z^1_1 = 0.0911 \cdot 0.0025 = 2.2775e - 04$ this value is cached
$\frac{\partial L}{\partial w^2_{22}} = (y - y*)w^3_{21} z^2_2 (1 - z^2_2) z^1_2 = 0.0911 \cdot 0.9975 = 0.0909$ this value is cached
Layer 1 weights:
$\frac{\partial L}{\partial w^1_{01}} = \frac{\partial L}{\partial y} \cdot (\frac{\partial y}{\partial z^2_1} \frac{\partial z^2_1}{\partial z^1_1} + \frac{\partial y}{\partial z^2_2} \frac{\partial z^2_2}{\partial z^1_1}) \cdot \frac{\partial z^1_1}{\partial w^1_{01}}$
$= (\frac{\partial L}{\partial z^2_1} z^2_1 (1 - z^2_1) w^2_{11} + \frac{\partial L}{\partial z^2_2} z^2_2 (1 - z^2_2) w^2_{12}) z^1_1 (1 - z^1_1) x_0$ this value is cached
$= ((y - y*)w^3_{11} z^2_1 (1 - z^2_1) w^2_{11} + (y - y*)w^3_{21} z^2_2 (1 - z^2_2) w^2_{12}) z^1_1 (1 - z^1_1) x_0$ this value is cached
$= (-0.1215 \cdot -2 + 0.0911 \cdot 2) \cdot 0.0025 \cdot (1 - 0.0025) \cdot x_0$ this value is cached
$= 0.0011$
$\frac{\partial L}{\partial w^1_{11}} = 0.0011 x_1 = 0.0011$
$\frac{\partial L}{\partial w^1_{21}} = 0.0011 x_2 = 0.0011$
$\frac{\partial L}{\partial w^1_{02}} = \frac{\partial L}{\partial y} \cdot (\frac{\partial y}{\partial z^2_1} \frac{\partial z^2_1}{\partial z^1_1} + \frac{\partial y}{\partial z^2_2} \frac{\partial z^2_2}{\partial z^1_1}) \cdot \frac{\partial z^1_1}{\partial w^1_{01}}$
$= (\frac{\partial L}{\partial z^2_1} z^2_1 (1 - z^2_1) w^2_{11} + \frac{\partial L}{\partial z^2_2} z^2_2 (1 - z^2_2) w^2_{12}) z^1_1 (1 - z^1_1) x_0$ this value is cached
$= ((y - y*)w^3_{11} z^2_1 (1 - z^2_1) w^2_{11} + (y - y*)w^3_{21} z^2_2 (1 - z^2_2) w^2_{12}) z^1_1 (1 - z^1_1) x_0$ this value is cached
$= (-0.1215 \cdot -3 + 0.0911 \cdot 3) \cdot 0.9975 \cdot (1 - 0.9975) \cdot x_0$ this value is cached
$= 0.0016$
$\frac{\partial L}{\partial w^1_{12}} = 0.0016 x_1 = 0.0016$
$\frac{\partial L}{\partial w^1_{22}} = 0.0016 x_2 = 0.0016$

4.

- MAP objective function: $max_w P(S|w)p(w)$ where $S(m = 3, d = 4)$ and $x_0 = 1$
  weights:
  $logP(S|w) = log(\prod_{i=1}^d p(y_i|x_i, w)) = log(\prod_{i=1}^m \frac{1}{1+e^{-y_i w^T x_i}}) = -\sum_{i=1}^d log(1+e^{-y_i w^T x_i})$

  $logp(w) = log(\prod_{i=1}^d p(w_i)) = log(\prod_{i=1}^d \frac{1}{2\sqrt{2\pi}} e^{-\frac{w_i^2}{2}}) = -\sum_{i=1}^d \frac{w_i^2}{2} - d \cdot log(2\pi) = -\frac{1}{2} w^T w + C$
  with the found weights the object function becomes:
  $min_w L(w) = \frac{1}{2} w^T w + \sum_{i=1}^m log(1 + e^{-y_i w^T x_i})$
  Gradient of the objective function:
  $\nabla L(w) = w - \sum_{i=1}^m y_i x_i (1 - \sigma(y_i w^T x_i))$
  $\nabla L(w, x_i, y_i) = w - m y_i x_i (1 - \sigma(y_i w^T x_i))$

- $y_1 w_0^T x_1 = 0$
  $\nabla L(w_0) = w_0 - 3(1-\sigma(0))[1, 0.5, -1, 0.3]^T = [-1.5, -0.75, 1.5, -0.45]^T$
  $w_1 = w_0 - 0.01\nabla L(w_0) = 0.01 \cdot [1.5, 0.75, -1.5, 0.45]^T$
  $y_2 w_1^T x_2 = -0.0285$
  $\nabla L(w_1) = w_1 - 3(1-\sigma(-0.0285))[1, -1, -2, -2]^T = [-1.5064, 1.5289, 3.0277, 3.0472]^T$
  $w_2 = w_1 - 0.005\nabla L(w_1) = 0.01 \cdot [1.5, 0.75, -1.5, 0.45]^T - 0.005 \cdot [-1.5064, 1.5289, 3.0277, 3.0472]^T = [0.0225, -0.0001, -0.0301, -0.107]^T$
  $y_3 w_2^T x_3 = 0.0431$
  $\nabla L(w_2) = w_2 - 3(1-\sigma(0.0431))[1, 1.5, 0.2, -2.5]^T = [-1.4452, -2.2016, -0.3236, 3.6585]^T$
  $w_3 = w_2 - 0.0025\nabla L(w_2)$
  $= [0.0225, -0.0001, -0.0301, -0.0107]^T - 0.0025 \cdot [-1.4452, -2.2016, -0.3236, 3.6585]^T = [0.0261, 0.0054, -0.0293, -0.0198]^T$

2 Practice [62 points + 50 bonus]

1. Github Link: https://github.com/BritGaul/CS5350

2. (a) Table of Variance and Errors:

Variance	Training Error	Testing Error
0.01	0.0401	0.0540
0.1	0.0115	0.0120
0.5	0.0115	0.0100
1	0.0115	0.0120
3	0.0115	0.0120
5	0.0161	0.0160
10	0.0115	0.0200
100	0.0298	0.0340

(b)

Table of Variance and Errors:

Variance	Training Error	Testing Error
0.01	0.0321	0.0400
0.1	0.0183	0.032
0.5	0.03444	0.0500
1	0.0103	0.0140
3	0.0264	0.0360
5	0.0161	0.0280
10	0.0344	0.0420
100	0.0057	0.0120

(c)

The ML estimation approach resulted in higher errors for the majority of the variances when compared to the MAP method. However, one exception was for the case of the variance being set to 100. With this variance the ML estimation method had smaller errors for both training and testing.

3.

(a) Code is in the Neural Networks folder in the file named backpropagation.py on github

(b)

Table of Width and Errors:

Width	Training Error	Testing Error
5	0.4461	0.4420
10	0.0046	0.0020
25	0.0482	0.0760
50	0.0069	0.0080
100	0.0321	0.0320

(c)

Table of Width and Errors:

Width	Training Error	Testing Error
5	0.4461	0.4420
10	0.4461	0.4420
25	0.4461	0.4420
50	0.4461	0.44200
100	0.4461	0.4420

The training and testing errors for all widths remain the same. When the weights are set to zero the neural network does not seem to improve with a chnage in width. This could be because the weights are not updated properly when initialized to zero.

(d)

The neural network and logistic regression seemed to generally perform better then the SVM. However for the neural network to perfrom better the weights need to be initialized properly, not to zero, and the width of the tree seemed to perform best when set to ten.

4.

It has been a fun class and I like how my machine learning library turned out. I have enjoyed this class and have learned a lot.