CS 613 - Machine Learning

Assignment 2 - Logistic Regression
Robert Thompson

1 Theory

1. For the function $J = (x_1 w_1 - 5x_2 w_2 - 2)^2$, where $w = [w_1, w_2]$ are our weights to learn:

   (a) What are the partial gradients, $\frac{\partial J}{\partial w_1}$ and $\frac{\partial J}{\partial w_2}$?

   i. Partial Gradient of $\frac{\partial J}{\partial w_1}$:

   A. Chain Rule
   $\frac{\partial J}{\partial w_1} = 2(x_1 w_1 - 5x_2 w_2 - 2)\ \frac{\partial J}{\partial w_1}\ (x_{1w1} - 5x_{2w2} - 2)$

   B. Sum/Difference Rule
   $\frac{\partial J}{\partial w_1}(x_{1w1} - 5x_{2w2} - 2) = \frac{\partial J}{\partial w_1}\ (x_1 w_1)$ - $\frac{\partial J}{\partial w_1}(5x_2 w_2)$ - $\frac{\partial J}{\partial w_1}(2)$
   $\frac{\partial J}{\partial w_1}(x_1 w_1) = x_1$
   $\frac{\partial J}{\partial w_1}(5x_2 w_2) = 0$
   $\frac{\partial J}{\partial w_1}(2) = 0$
   $= x_1 - 0 - 0$
   $= x_1$

   C. Partial Gradient $\frac{\partial J}{\partial w_1} = 2x_1(x_1 w_1 - 5x_2 w_2 - 2)$

   i. Partial Gradient of $\frac{\partial J}{\partial w_2}$:

   A. Chain Rule
   $\frac{\partial J}{\partial w_2} = 2(x_1 w_1 - 5x_2 w_2 - 2)\ \frac{\partial J}{\partial w_2}\ (x_{1w1} - 5x_{2w2} - 2)$

   B. Sum/Difference Rule
   $\frac{\partial J}{\partial w_2}(x_{1w1} - 5x_{2w2} - 2) = \frac{\partial J}{\partial w_2}\ (x_1 w_1)$ - $\frac{\partial J}{\partial w_2}(5x_2 w_2)$ - $\frac{\partial J}{\partial w_2}(2)$
   $\frac{\partial J}{\partial w_2}(x_1 w_1) = 0$
   $\frac{\partial J}{\partial w_2}(5x_2 w_2) = 5x_2$
   $\frac{\partial J}{\partial w_2}(2) = 0$
   $= 0 - 5x_2 - 0$
   $= -5x_2$

   C. Partial Gradient $\frac{\partial J}{\partial w_2} = -10x_2(x_1 w_1 - 5x_2 w_2 - 2)$

   (b) What are the value of the partial gradients given current values of $w = [0, 0], x = [1, 1]$?

   i. Let $\frac{\partial J}{\partial w_1} = 2x_1(x_1 w_1 - 5x_2 w_2 - 2)$

   A. Plugin $w = [0, 0], x = [1, 1]$
   $\frac{\partial J}{\partial w_1} = 2(1)(1 * 0 - 5(1)(0) - 2)$

   B. Partial Gradient Value of $\frac{\partial J}{\partial w_1} = -4$

   ii. Let $\frac{\partial J}{\partial w_2} = -10x_2(x_1 w_1 - 5x_2 w_2 - 2)$

A. Plugin $w = [0, 0], x = [1, 1]$
$\frac{\partial J}{\partial w_2} = -10(1)(1 * 0 - 5(1)(0) - 2)$

B. Partial Gradient Value of $\frac{\partial J}{\partial w_2} = 20$

2 Spambase Logistic Regression Classier

1. Plot of Training and Validation Log Loss as a Function of the Epoch
    - Learning Rate: 0.1
    - Epochs: $10,000$
    - Stability Constant: $10e-7$

2. Training Statistics
    - Precision: 0.9267676767676768
    - Recall: 0.8900565885206144
    - F-Measure: 0.9080412371134021

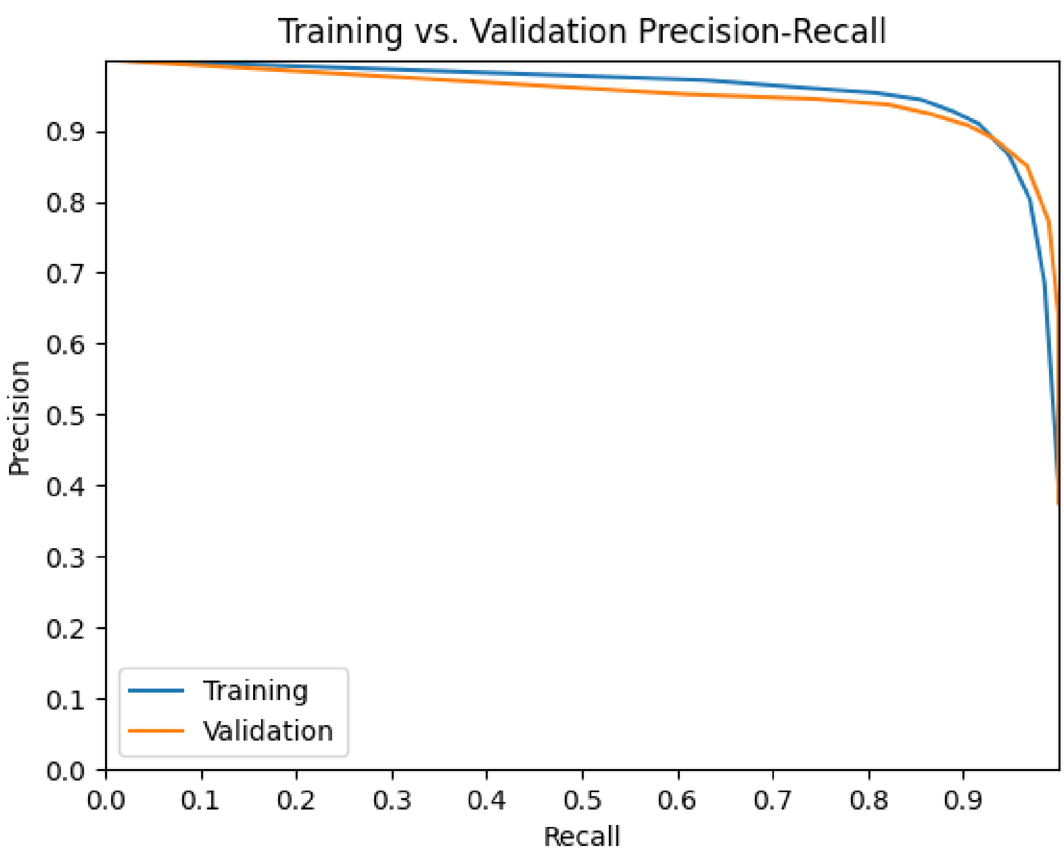
- Accuracy: 0.9272905119008803

3. Validation Statistics

- Precision: 0.9076655052264808
- Recall: 0.9045138888888888
- F-Measure: 0.9060869565217391
- Accuracy: 0.9295958279009127

4. Training and Validation Precision-Recall Graph

- Learning Rate: 0.1
- Epochs: $10,000$
- Stability Constant: $10e-7$



3 Logistic Regression for Multi-Class Classification

1. Validation Accuracy: 0.84

2. Validation Confusion Matrix

   - $\begin{bmatrix} 19 & 0 & 0 \\ 0 & 9 & 2 \\ 0 & 0 & 14 \end{bmatrix}$