Theodore Kim

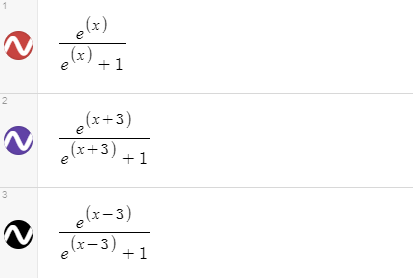
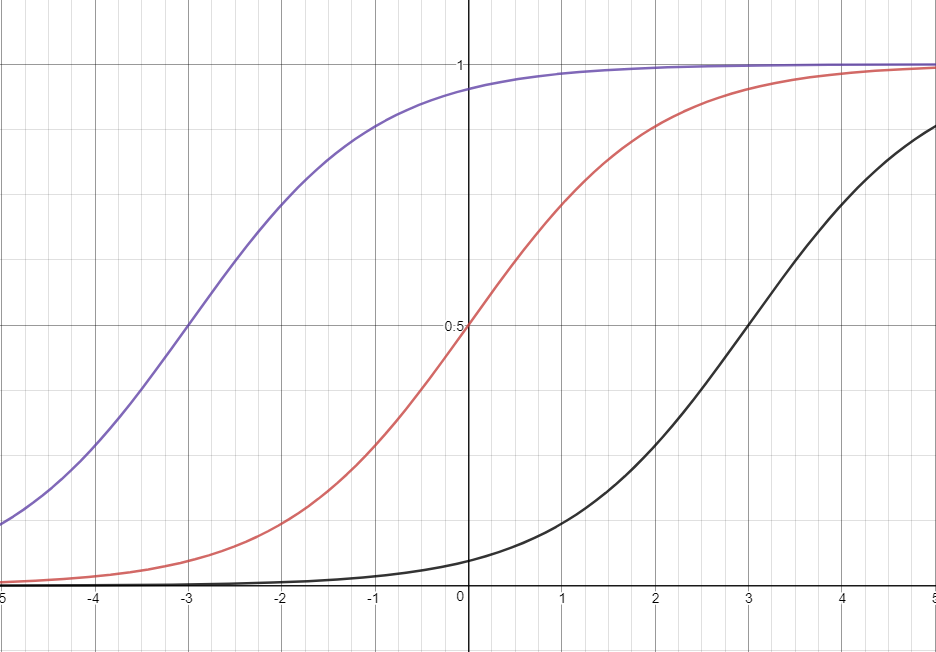
CS 4563-A: Introduction to Machine Learning

Homework #4

*1. How does the logistic function change when w0 changes?*

Changing w0 would cause the delimiting line (i.e. the line that separates the two populations) of the classifier to shift either upwards or downwards on the graph: if w­0 is increased the line would shift upwards, whereas the opposite is true of reducing w0.

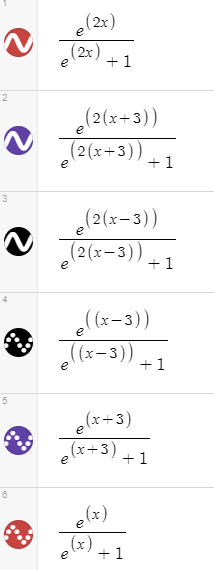
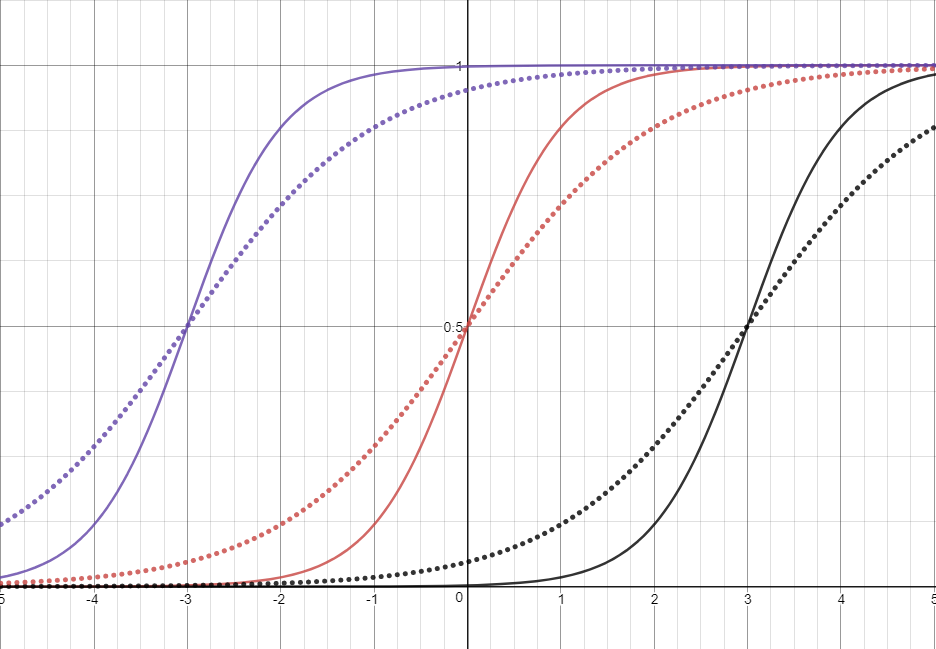
This shift in the line affects the logistic function as it affects the probability of the classification prediction. If the delimiter line is shifted upwards, there exists more "sample space" or possible samples that would be classified under the line and less samples that could possibly be classified above the line. Therefore, the "median" of the logistic function (i.e. the value of x at which the probability of prediction is 0.5 (i.e. the x value that lies ON the delimiter) would shift to the left if w0 was increased as the portion of the domain that falls below 0.5 would decrease. The opposite is true if w0 was decreased.



*A simulation of the logistic function with altered w0 values*

2. How does the logistic function change when if you use **w' =** **2w** instead of **w**?

Using **2w** rather than **w** would cause both the slope and intercept of the delimiter line to be doubled. Since both the slope and the intercept of the line is being altered by the same factor, the line is not translated in any direction (the altered lines will simply rotate by a fixed point). However, the line will become progressively more vertical, and therefore will result in a reduced "uncertainty" when choosing a prediction from a given x-value, in other words if a vertical line was used as the delimiter, there would be little uncertainty when determining "which side of the line" that a given x value belonged. Consequentially, the logistic function becomes more steep (there are less values of x that result in a y value between in the range [-0.9, 0.9]) and as there are less values of x that are not "extremely" far from the delimiter, therefore, not explicitly 1 or 0.



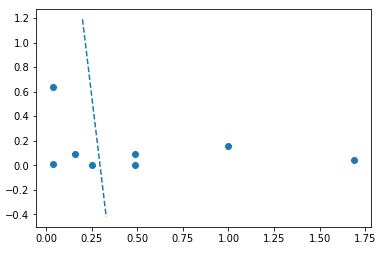
*A simulation of the logistic function using 2w instead of w*

3. Training a logistic classifier results in the following values for wT = [0.66, -0.24, -0.18]

a. Create a confusion matrix for a threshold of 0.5:

|  |  |  |  |
| --- | --- | --- | --- |
| n = 8 | **Predicted TRUE** | **Predicted FALSE** |  |
| **Actual TRUE** | 3 | 1 | 4 |
| **Actual FALSE** | 1 | 3 | 4 |
|  | 4 | 4 | 8 |

b. Graph the points and the delimiter:



**c** and **d**. Find the FPR and TPR

**e**, **f**, **g**. Calculate accuracy, precision and recall

**h**. How likely is the **w** for the examples above?

**i.** Given **w** as described above and **w'** = (1.33, -2.96, -2.77), which one is more likely to be the correct decision boundary given access to the data above.

**w'** is more likely to the be the correct decision boundary.

**j.** Perform one step of gradient descent using the **w** above. **Assuming learning rate is 0.1**

**k.** How did the data points near the decision boundary contribute to the new value of w?

***l****. How did the data points which were correctly classified and far away from the boundary contribute to the new value of w?*

These points influenced the values of w1 and w2 very little as their derivatives were closely clustered together (their sums were half of that of the sums of the incorrectly classified points).

***m****. How did incorrectly classified points contribute to the new value of w?*

These points did not influence the value of w0 very significantly. While the difference between y and yhat was the greatest for these points, the two incorrect points were equal in absolute value, buit had the opposite sign, so when summed, cancelled each other out. However, they did greatly influence the values of w1 and w2 as the difference between the two incorrect points was the greatest compared to the difference between the correctly classified point. For w1 the false positive was much greater than the false negative, where as for w2 the inverse was true. Given that the correctly classified points did not have much influence on these features of w, the incorrectly classified points were the most significant.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 0.49 | 0.09 | 0.389 | 0 | -0.389 | -0.19061 | -0.03501 |
| 1.69 | 0.04 | 0.042 | 0 | -0.042 | -0.07098 | -0.00168 |
| 0.04 | 0.64 | 0.613 | 0 | -0.613 | -0.02452 | -0.39232 |
| 1 | 0.16 | 0.167 | 0 | -0.167 | -0.167 | -0.02672 |
| 0.16 | 0.09 | 0.572 | 1 | 0.428 | 0.06848 | 0.03852 |
| 0.25 | 0 | 0.526 | 1 | 0.474 | 0.1185 | 0 |
| 0.49 | 0 | 0.393 | 1 | 0.607 | 0.29743 | 0 |
| 0.04 | 0.01 | 0.638 | 1 | 0.362 | 0.01448 | 0.00362 |