Introduction to Machine Learning
Homework 4: Logistic Regression*

Prof. Linda Sellie

1. How does the logistic function change when w_0 changes?
2. How does the logistic function change if you use $\mathbf{w}' = 2\mathbf{w}$ instead of $\mathbf{w}$?
3. Suppose you trained a logistic classifier on a data set and it produced coefficients: $\mathbf{w}^T = [0.66, -2.24, -0.18]$.

   In the table below $h_{\mathbf{w}}(x) = \frac{1}{1+e^{-w_0+w_{1:k}^T x}}$

   For the following examples:

	x_1	x_2	$h_\beta(x)$	y
1	0.49	0.09	0.389	0
2	1.69	0.04	0.042	0
3	0.04	0.64	0.613	0
4	1.	0.16	0.167	0
5	0.16	0.09	0.572	1
6	0.25	0.	0.526	1
7	0.49	0.	0.393	1
8	0.04	0.01	0.638	1

   (a) For a decision boundary of 0.5 create the confusion matrix.

   (b) Plot the points on a graph and draw the decision boundary (I would suggest using some sort of plotting library and a image editor)

   (c) For the data set above what is the FPR?

   (d) For the data set above what is the TPR?

   (e) What is the accuracy?

   (f) What is the recall?

   (g) What is the precision

   (h) How likely is the $\mathbf{w}$ for the examples above

   (i) Given $\mathbf{w}$ as described above and $\mathbf{w}' = (1.33, -2.96, -2.77)$, which is more likely to be the correct decision boundary given access only to the data above.

   (j) Perform one step of gradient ascent using the $\mathbf{w}$ above.

*Most of these questions are from Prof. Rangan.

(k) How did the data points near the decision boundary contribute to the new value of $\mathbf{w}$?

(l) How did the data points which were correctly classified and far away from the decision boundary contribute to the new value of $\mathbf{w}$?

(m) How did incorrectly classified points contribute to the new value of β?

4. Regularization:

   - Add lasso regularization to the log likelihood function for logistic regression
   - Add ridge regularization the log likelihood function for logistic regression
   - Determine the derivative of log likelihood function for logistic regression with ridge regularization. Then:
     - Modify your lab for logistic regression to now include ridge regularization.
     - Add to this lab 5-fold cross validation to find the optimal λ value and answer the following questions. Did the regularization help? How did the regularization affect the training data. How did it affect the validation set?
     - Add to this lab by writing the code to calculate the confusion matrix for the test data. Calculate the values of the confusion matrix for the test data using the best coefficient vector you found
     - Turn in your python notebook for this lab

5. Suggest possible response variables and predictors for the following classification problems. For each problem, indicate how many classes there are. There is no single correct answer.

   (a) Given an audio sample, to detect the gender of the voice.

   (b) A electronic writing pad records motion of a stylus and it is desired to determine which letter or number was written. Assume a segmentation algorithm is already run which indicates very reliably the beginning and end time of the writing of each character.

6. A data scientist is hired by a political candidate to predict who will donate money. The data scientist decides to use two predictors for each possible donor:

   - x_1 = the income of the person(in thousands of dollars), and
   - x_2 = the number of websites with similar political views as the candidate the person follows on Facebook.

   To train the model, the scientist tries to solicit donations from a randomly selected subset of people and records who donates or not. She obtains the following data:

Income (thousands \$), x_{i1}	30	50	70	80	100
Num websites, x_{i2}	0	1	1	2	1
Donate (1=yes or 0=no), y_i	0	1	0	1	1

   (a) Draw a scatter plot of the data labeling the two classes with different markers.

(b) Find a linear classifier that makes at most one error on the training data. The classifier should be of the form,

$$\hat{y}_i = \begin{cases} 1 & \text{if } z_i > 0 \\ 0 & \text{if } z_i < 0, \end{cases} \qquad z_i = \boldsymbol{\beta}_{1:k}^{\mathsf{T}} \mathbf{x}_i + b.$$

What is the weight vector $\mathbf{w}$ of your classifier?

(c) Now consider a logistic model of the form,

$$P(y_i = 1|\mathbf{x}_i) = \frac{1}{1 + e^{-z_i}}, \qquad z_i = \mathbf{w}1 : k^{\mathsf{T}} \mathbf{x}_i + w_0.$$

Using $\mathbf{w}$ from the previous part, which sample i is the *least* likely (i.e. $P(y_i|\mathbf{x}_i)$ is the smallest). If you do the calculations correctly, you should not need a calculator.

(d) Now consider a new set of parameters

$$\mathbf{w}' = \alpha \mathbf{w},$$

where $\alpha > 0$ is a positive scalar. Would using the new parameters change the values $\hat{y}$ in part (b)? Would they change the likelihoods $P(y_i|\mathbf{x}_i)$ in part (c)? If they do not change, state why. If they do change, qualitatively describe the change as a function of α.