# Logistic Regression Exercises

The following example problems, and their solutions, illustrate the material covered in class (and in the Lecture 4 slides) on logistic regression.

1. Suppose we train a logistic classifier to solve a binary classification problem (i.e. we do logistic regression). Assume that we train the classifier using gradient descent, as described in lecture and in the textbook. At each stage of the gradient descent, values are assigned to the weights of the classifier. The classifier corresponds to a function of the form $h(x) = \frac{1}{1+e^{-(w^T x+w_0)}}$ whose output is an estimate of $P(C_1|x)$. It classifies an example as belonging to $C_1$ if $h(x) > 1/2$, and to $C_2$ otherwise.

   Assume that in the classification problem, there are two real-valued attributes, $x_1$ and $x_2$, and the training set contains only two examples, $x^{(1)} = (x_1 = -3, x_2 = 4)$ with label $r = 1$, and $x^{(2)} = (x_1 = 2, x_2 = 3)$ with label $r = 0$. (Here $r = 1$ denotes that $x \in C_1$, and $r = 0$ denotes that $x \in C_2$). At the start of the gradient descent procedure, let the initial weights be $w_0 = w_1 = w_2 = 0.01$. Assume we used a fixed learning rate of $\eta = .005$.

   (a) What is the predicted (estimated) value of $P(C_1|x^{(1)})$, using the initial weights?

   *It is* $.505$*, because it is equal to* $1/(1+e^{-g(x)})$ *where* $g(x) = .01 * -3 + .01 * 4 + .01 = 0.2$

   (b) What is the predicted *label* of $x^{(1)}$, using the initial weights. (Recall that we label an example as belonging to $C_1$ if $P(C_1|x) > 1/2$.) Is the prediction consistent with the label of this example in the training set?

   *Predicted label is 1, for class* $C_1$*. This is consistent with the label in the training set for* $x^{(1)}$*.*

   *Note that it's not actually necessary to compute* $1/(1 + e^{-(w^T x+w_0)})$ *if you just want to predict the label for* $x^{(1)}$*. It's enough to compute the linear discriminant function* $g(x) = .01 * -3 + .01 * 4 + .01 = 0.2$*, and to check whether it is greater than 0.*

   (c) Repeat the above two questions for $x^{(2)}$.

   *The predicted value is* $.515$*. Predicted label is 1. This is not consistent with label in training set.*

   (d) Again using the initial weights, what is the training error of the classifier? That is, what percentage of the labels in the training set does it predict incorrectly?

   *50% error*

   (e) In logistic discrimination, we are trying to minimizing the error function $Err(\mathbf{w}, \mathbf{w_0}|\mathcal{X}) = -\sum_t (r^t \log y^t + (1 - r^t) \log(1 - y^t))$. This quantity is sometimes called the *cross-entropy* of the classifier on the dataset. Using the initial weights, what is the cross-entropy of the classifier on the given training set?

$-(log(1/(1+e^{-(.01*-3+.01*4+.01)})+log(1-(1/(1+e^{-(.01*2+.01*3+.01)})))) = 1.4068$

(f) Suppose we do one iteration of the gradient descent procedure (i.e., one iteration of the main loop, so we update each weight once) on this training set, beginning with the initial weights. Give the values of the updated weights.

$w_1 = -0.0025750, w_2 = 0.0121751, w_0 = 0.0099000$