(1) This is an open book, take home quiz. **No collaboration is allowed.**

(2) Remember that your work is graded on the quality of your writing and explanation as well as the validity of the mathematics.

(1) Sometimes machine learning is used on imperfect training data – for example, data collected via noisy sensors. In these cases, we might try to correct for noise while training the classifier.

Consider the following formulation for training a logistic regression classifier $w \in \mathbb{R}^d$ on a noisy training data set $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})$, where for each $i$, $y^{(i)} \in \{-1, +1\}$. For simplicity, we ignore the bias term $b$. Suppose we know that the noise magnitude is at most $r$. Then, instead of the standard logistic regression loss, we might want to minimize the following loss: 

$$ \tilde{L}(w) = \sum_{i=1}^n \tilde{L}_i(w), \text{ where,} $$

$$ \tilde{L}_i(w) = \max_{z: \|z\| \leq r} \log(1 + \exp(-y^{(i)} w^\top z^{(i)})),$$

where $\|v\|$ means the $L_2$-norm of vector $v$.

(a) (5 points) Prove that for all $i$, $\tilde{L}_i(w) = M_i(w)$, where $M_i(w) = \log(1 + \exp(r \|w\| - y^{(i)} w^\top x^{(i)}))$. For full credit, show all the steps in your proof.

(b) (5 points) Let $M(w) = \sum_{i=1}^n M_i(w)$. Write down the gradient descent update for minimizing $M(w)$. 


(c) (5 points) Suppose you already have code for a single stochastic gradient update for minimizing the logistic regression loss function. Specifically, you have code for a function \texttt{logistic-SGD}(w, x^{(i)}, y^{(i)}, \eta) that given the current \(w\), a labeled example \((x^{(i)}, y^{(i)})\) and a learning rate \(\eta\), returns you the updated weight vector:

\[ w' = w - \eta \nabla_w \log(1 + e^{-y^{(i)} w^\top x^{(i)}}) \]

Write down how you can use this function \texttt{logistic-SGD} to code up a stochastic gradient update on \(\hat{L}_i(w) = M_i(w)\). Specifically, given a \(w\), a labeled example \((x^{(i)}, y^{(i)})\) and a learning rate \(\eta\), your function should return the updated weight vector:

\[ w'' = w - \eta \nabla_w M_i(w) \]

(Hint: You may need to give \texttt{logistic-SGD} inputs that are different from \(w, (x^{(i)}, y^{(i)})\) and \(\eta\).)