CSE 251A Winter 2021 - Take Home Exam 2

1 hour **Name:** **PID:**

---

(1) This is an open book, open notes exam. You are free to consult any text book or notes. **You are not allowed to consult with any other person.**

(2) If you need any clarification, please post a private message to the instructors on Piazza.

(3) Remember that your work is graded on the *clarity* of your writing and explanation as well as the validity of what you write.

(4) This is a one-hour exam.

(1) Suppose we are given a training set $S = \{(x^{(i)}, y^{(i)}), i = 1 \dots n\}$ of labeled vectors where $x^{(i)} \in \mathbb{R}^d$ are d-dimensional feature vectors and $y^{(i)} \in \{-1, 1\}$ are labels. Let (w_{LR}, b_{LR}) be the solution to L_2-regularized logistic regression with regularization parameter λ on S.

Now suppose that we scale each feature vector by a positive constant $c > 1$ to get the training dataset S' – that is, $S' = \{(cx^{(i)}, y^{(i)}), i = 1, \dots, n\}$. Let (w'_{LR}, b'_{LR}) be the solution to L_2-regularized logistic regression with regularization parameter λ' on S'.

State whether the following statements are true or false. Justify your answer with either a proof or a counterexample.

(a) (5 points) Suppose that $\lambda' = \lambda = 0$, and that (w_{LR}, b_{LR}) and (w'_{LR}, b'_{LR}) are unique. Then $w'_{LR} \neq w_{LR}$ in general (for any arbitrary S).

(b) (5 points) If $\lambda' = c^2\lambda > 0$, then $w'_{LR} = w_{LR}/c$ and $b'_{LR} = b_{LR}$.

(2) Suppose we are given an input dataset $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ where each $x^{(i)} \in \mathbb{R}^d$. Let v be a parameter vector in $\mathbb{R}^d$. Consider the following loss function:

$$G(v) = \sum_{i=1}^{n} \frac{v^\top x^{(i)} (x^{(i)})^\top v}{v^\top v}$$

(a) (5 points) Write down the update step for stochastic gradient descent corresponding to a single data point $x^{(i)}$.

(b) (5 points) Write down an algorithm that implements the stochastic gradient update in the first part in $O(d)$ time.