APMA E4990.002: QUIZ 2
DUE TUE, NOV 22 AT 11:59PM ET

VLADIMIR KOBZAR

Please justify your answers, proving the statements you make. You are allowed to refer to results shown in lecture (or that are in the textbook) as long as you state them precisely, meaning that you should say exactly which hypothesis are needed in the result you use.

This quiz is open book, but you are not allowed to share or discuss this quiz with any person (including not posting regarding this quiz on Ed discussion or elsewhere). If you have any questions or find errors please email me at vak2116@columbia.edu.

If you need to impose extra conditions on a problem to make it easier (or consider specific cases of the question, like taking n to be 2, e.g.), state explicitly that you have done so. Solutions where extra conditions were assumed, or where only special cases where treated, will also be graded (probably scored as a partial answer).

Note that it may take some time to upload the quiz on Gradescope. Please plan accordingly. I suggest uploading your quiz at least 30 minutes before the deadline. Make sure your answers to each problem are clearly stated in the submitted PDF. All code (Jupyter Notebooks and other source files) used to compute your answers should also be uploaded to Gradescope, along with a readme file telling the graders how to run your code if they need to, and if you are using any special packages. Include a .zip file with your online homework submission containing all of your source code.

(1) Given a set of data points $(x_1, y_1), ..., (x_n, y_n)$ where the features $x_i \in \mathbb{R}^m$ and binary labels $y_i \in \{0, 1\}$ determine whether gradient descent convergences to a global minimum or a local critical point when you minimize the cross-entropy loss

$$L(w) = -\frac{1}{n}\sum_{i=1}^{n} y_i \log f(w^T x_i) + (1 - y_i)\log(1 - f(w^T x_i))$$

where $f : \mathbb{R} \to (0, 1)$ is defined by

$$f(t) = \frac{1}{1 + e^{-t}}$$

(the *logistic regression* model) and determine the convergence rate of the gradient descent. You can use any of the methods we discussed in class for determining the step size (but note that backtracking line search is recommended for the next problem, so you may consider using it in your convergence analysis here). How would adding an ℓ_2 regularization term to $L(w)$ affect the convergence analysis?

(2) The code you will implement in this question is located in the `logistic.py` file in the `mnist` folder of `quiz2.zip`. In this problem we will train and test a simple logistic regression classifier that will determine whether a given image of a digit represents a 5 or not. More

Date: Nov 15, 2022.

precisely, given a dataset $(x_1, y_1), ..., (x_n, y_n) \in \mathbb{R}^{784} \times \{0, 1\}$ we will find $w \in \mathbb{R}^{784}$ that minimizes the empirical loss function L as defined in the previous problem. Here $y_i = 1$ when the corresponding image x_i is a 5, and $y_i = 0$ otherwise.

(a) Implement the L function in `logistic.py`. [Hint: Note that $\log(1 - f(w^T x_i))$ and $\log f(w^T x_i)$ are both logs of exponentials, and thus should be well-behaved even when $w^T x_i$ is very small or very large. Can evaluate with logsumexp or something like it for greater numerical precision (i.e., you shouldn't get $\pm\infty$). Also note that $1 - f(t) = e^{-t}/(1 + e{-}t)$.]

(b) Compute the gradient $\nabla L(w)$ of L with respect to w.

(c) Using the previous part, implement the dL function in `logistic.py`.

(d) Implement the gradient descent function which minimizes L using gradient descent with the step size method that you used in the previous problem (backtracking line search is recommended). See the comments for more details. Include the training loss and test errors output by the code in your submission, along with the generated plots.

(e) Suppose your code terminates after some number of iterations. How does the theoretical convergence guarantee determined in the previous problem compare with what the code actually computes?

(3) In this exercise you will use the code in the `mnist.ipynb` in the `mnist_nn` folder of `quiz3.zip`. Please provide code responsive to questions (a)-(c) described in that notebook. I suggest starting by running the cells in the given notebook corresponding to question (b), which should be already functional, to make sure that the notebook runs properly. Include the .pdf of the notebook with responses to questions 1 and 2 above, and submit the notebook separately in .ipynb format.

Acknowledgement: Problems in this homework set are based on materials in courses taught by Dr. Brett Bernstein and Prof. Carlos Fernandez-Granda. These materials are used with the permission of the authors.