

## Homework 2

Please upload your assignments on or before March 2, 2021.

- You are encouraged to discuss ideas with each other; but
  - you **must acknowledge** your collaborator, and
  - you **must compose your own** writeup and/or code independently.
  - We **require** answers to theory questions to be written in LaTeX, and answers to coding questions in Python (Jupyter notebooks)
  - Upload your answers in the form of a single PDF on Gradescope.
- 

1. **(4 points)** Analyzing gradient descent. Consider a simple function having two weight variables:

$$L(w_1, w_2) = 0.5(aw_1^2 + bw_2^2).$$

- a. Write down the gradient  $\nabla L(w)$ , and using this, derive the weights  $w^*$  that achieve the minimum value of  $L$ .
- b. Instead of simply writing down the optimal weights, let's now try to optimize  $L$  using gradient descent. Starting from some randomly chosen (non-zero) initialization point  $w_1(0), w_2(0)$ , write down the gradient descent updates. Show that the updates have the form:

$$w_1(t+1) = \rho_1 w_1(t), \quad w_2(t+1) = \rho_2 w_2(t)$$

where  $w_i(t)$  represent the weights at the  $t^{\text{th}}$  iteration. Derive the expressions for  $\rho_1$  and  $\rho_2$  in terms of  $a, b$ , and the learning rate.

- c. Under what values of the learning rate does gradient descent converge to the correct minimum? Under what values does it not?
  - d. Provide a scenario under which the convergence rate of gradient descent is very slow. (*Hint: consider the case where  $a/b$  is a very large ratio.*)
2. **(6 points)** Open the (incomplete) Jupyter notebook provided as an attachment to this homework in Google Colab (or other cloud service of your choice) and complete the missing items. Save your finished notebook in PDF format and upload along with your answers to the above theory questions in a single PDF.