Assignment V

MEMAD

September 13, 2025

Abstract

The goal is to revise some relevant aspects of stochastic optimization models, especially the minibatch gradient descent and Adam methods. Please upload your solutions in one compressed file to Classroom before September 23rd.

Problem 1

Consider the following function

$$f(x) = 2^{cos(x^2)}, \qquad x \in (-\pi, \pi).$$

Next, do the following:

- Compute 500 pair sample data $D = \{x_i, f(x_i)\}$ where the x_i are equispaced.
- Make a plot of D.
- Consider a polynomial model of degree $\mathcal{N}$:

$$h(x_i; \theta) = \hat{y}_i = \sum_{\ell=0}^{\mathcal{N}} \theta_\ell x_i^\ell, \qquad \theta = (\theta_0, \theta_1, \theta_2, \ldots, \theta_{\mathcal{N}}) \in \mathbb{R}^{\mathcal{N}+1}.$$

- Consider a MSE loss and cost functions:

$$Loss(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2,$$

$$Cost(\theta; D) = \frac{1}{m} \sum_{i=1}^{m} Loss(y_i, \hat{y}_i).$$

- Code a Python script to solve the associated optimization problem using some deterministic line search algorithm. You can try out $\mathcal{N} = 8$ for instance.
- Make a plot of the solution against $f(x_i)$ and show in another plot how the cost decreases as the iterations go by.

Problem 2

Attempt to find a better size for the parametric model in the previous problem by solving again for several values of $\mathcal{N}$. Make some plots showing your findings. From herein you can fix this value of $\mathcal{N}$ for the remaining problems.

Problem 3

Make a script to solve Problem 1 using a minibach gradient descent with a constant learning rate α. Make a plot of the solution against $f(x_i)$ and show in another plot how the cost decreases as the iterations go by.

Problem 4

Repeat problem 3 using Adam.

Problem 5

Make a comparison of the solutions and performance by the three algorithms you have used to solve problem 1. Write down your conclusions and make some plots.

Problem 6

So far, you have solved a noise-free problem, i.e.

$$f(x) = 2^{cos(x^2)} + \mathcal{X}, \qquad x \in (-\pi, \pi),\ \mathcal{X} \sim N(0, \sigma^2),$$

where you have only considered the case with $\sigma^2 = 0$. Repeat problems 1-5 for the noise levels $\sigma^2 = 0.05, 0.1, 0.5$. Don't forget to write down all your conclusions and thoughts.

Problem 7

Explain using your own words the following:

a) The differences and similarities between stochastic gradient descent, steepest descent, and minibatch gradient descent.

b) Some examples of the usage for the latter three cited algorithms (i.e., explain how to identify when an algorithm should be used among the others).

c) Some examples of the usage of a deterministic approach (Newton, steepest descent, BFGS, Gauss-Newton, etc) and a stochastic approach (stochastic gradient descent, minibatch gradient descent, Adam, etc), i.e., explain how to identify when a particular approach should be used on top of another.