CAS ETH Machine Learning in Finance and Insurance. Mini-exercises - Lecture 2.

Andrea Ferrario

Version of March 14, 2024

1 Gradients and contours of functions in many variables

1. Explain what a partial derivative of a function in multiple variables is and what the gradient of that function is.

2. Explain what a minimum point of a function is. Provide a graphical example of a function with (1) no minimum point, (2) one minimum point, (3) three minimum points, and (4) countably many minimum points. Show that, if a function has a minimum *value*, then it is unique. Also explain what a local minimum point of a function is.

3. Compute the gradient of the function $f(x, y) = 3x^2 + 2y - 5$, $(x, y) \in \mathbb{R}^2$.

4. Find the gradient of the function $g(x, y, z) = \sin(x) + \cos(y) + e^z$, $(x, y, z) \in \mathbb{R}^3$.

5. Find the domain of the function $h(x, y) = \log(x) + \log(y) + \log(x - y)$ and compute its gradient.

6. Determine at which points $(x, y, z) \in \mathbb{R}^3$ the function $h(x, y, z) = \sqrt{x^2 + y^2 + z^2}$ has a gradient.

7. Describe the contours of the function $f(x, y) = x^2 + y^2 - 1$, where $(x, y) \in \mathbb{R}^2$.

8. Describe the contours of the function $g(x, y) = xy - 1$, $(x, y) \in \mathbb{R}^2$.

9. Describe the contours of the function $h(x, y) = x^2 - y^2$, $(x, y) \in \mathbb{R}^2$.

10. Write Python code to compute and plot the gradient of the function $f(x, y) = 3x^2 + 2y - 5$, $(x, y) \in \mathbb{R}^2$.

11. Write Python code to visualize the contours of the function $f(x, y) = x^2 + y^2 - 1$, $(x, y) \in \mathbb{R}^2$. Let the user choose any $(x, y) \in \mathbb{R}^2$ and show the gradient $\nabla f(x, y)$ on the contour plot.

12. Write Python code to visualize the contours of the function $g(x, y) = xy - 1$, $(x, y) \in \mathbb{R}^2$. Let the user choose any $(x, y) \in \mathbb{R}^2$ and show the gradient $\nabla f(x, y)$ on the contour plot.

13. Write Python code to visualize the contours of the function $h(x, y) = x^2 - y^2$, $(x, y) \in \mathbb{R}^2$. Let the user choose any $(x, y) \in \mathbb{R}^2$ and show the gradient $\nabla f(x, y)$ on the contour plot.

2 (Stochastic) Gradient Descent

1. Consider the function $f(x) = x^2 - 4x + 5$, $x \in \mathbb{R}$. Perform gradient descent (GD) starting from $x_0 = 0$ with a learning rate of $\eta = 0.1$. Execute the optimization for 5 iterations, reporting the current iteration number, the current x value, and the corresponding function value at each iteration.

2. Consider the function $f(x) = \frac{1}{3}x^3 - x^2 + 2x$, $x \in \mathbb{R}$. Perform GD starting from $x_0 = 3$ with a learning rate of $\eta = 0.5$. Compute the sequence $x_0, x_1, \ldots, x_9$. What do you observe?

3. (Python) Implement GD to minimize the function $f(x) = x^3 - 2x^2 - 5x + 6$, $x \in \mathbb{R}$. Start from $x_0 = 2$ with a learning rate of $\eta = 0.1$. Execute the optimization until the absolute difference between consecutive iterations is less than 0.001. Report the final value of x and the minimum function value reached by the algorithm.

4. Explain the concepts of stochastic gradient descent (SGD) and mini-batch SGD and how they differ from GD. Evaluate the advantages and disadvantages of these SGD algorithms in comparison to GD.

5. (Python) Implement mini-batch SGD to minimize the function $E(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m} L(f_{\boldsymbol{\theta}}(x_i), y_i)$, where $L(f_{\boldsymbol{\theta}}(x_i), y_i)$ is the squared loss and $f_{\boldsymbol{\theta}}(x) = \theta_0 + \theta_1 x_1 + \cdots + \theta_d x_d$ is a linear regression model in $d > 1$ variables. Simulate N data points $\{(x_i, y)\}_{i=1}^{N}$, where $y_i = f_{\boldsymbol{\theta}}(x_i) + \epsilon_i$ (choose the errors appropriately). Initialize $\boldsymbol{\theta}$, choose different values of the learning rate and execute SGD for different numbers of iterations and mini-batch sizes. Comment on results.