Programming Assignment 1

CS6230: Optimization Methods in Machine Learning
IIT-Hyderabad
Aug-Nov 2017

Max Marks: 40
Due: 7th Sep 2017 11:59 pm

Instructions

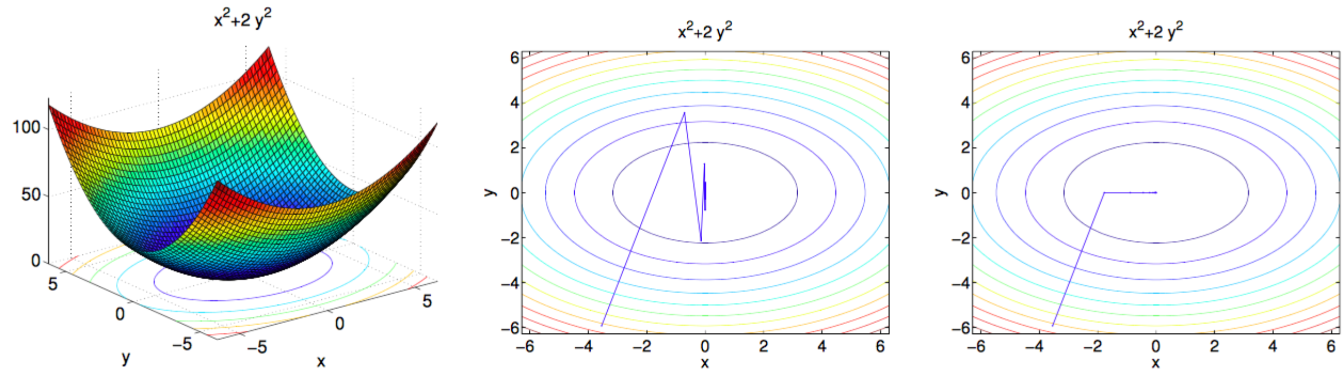
- Please use Google Classroom to upload your submission by the deadline mentioned above. Your submission should comprise of a single ZIP file, named `<Your_Roll_No>_PA1`, with all your solutions.
- For late submissions, 10% is deducted for each day (including weekend) late after an assignment is due. Note that each student begins the course with 6 grace days for late submission of assignments. Late submissions will automatically use your grace days balance, if you have any left. You can see your balance on the CS6230 Marks and Grace Days document under the course Google drive.
- You should use Python for this assignment.
- Please read the department plagiarism policy. Do not engage in any form of cheating - strict penalties will be imposed for both givers and takers. Please talk to instructor or TA if you have concerns.

1 Programming Assignment

For this assignment, consider the following four functions which we seek to minimize: a quadratic function, a ridge regularized logistic regression, the Himmelblaus function and Rosenbrock's banana function:

$$\begin{aligned}
f_Q(x,y) &= 1.125x^2 + 0.5xy + 0.75y^2 + 2x + 2y \\
f_{LL}(x,y) &= 0.5(x^2+y^2) + 50\cdot\log(1+exp(-0.5y)) + 50\cdot\log(1+exp(0.2x)) \\
f_H(x,y) &= 0.1(x^2+y-11)^2 + 0.1(x+y^2-7)^2 \\
f_R(x,y) &= 0.002(1-x)^2 + 0.2(y-x^2)^2
\end{aligned}$$

Answer the following questions:



(a) function visualization (b) fixed stepsize 0.4 (c) backtracking

Figure 1: Sample illustrations

1. *(6 marks)* Plot and visualize the four functions (similar to Figure 1a). The first three functions are on the domain $[-6,6][-6,6]$ and the last f_R is on $[-3,3][-6,6]$. Which functions are convex? Which are not?

2. *(10 marks)* For each of the functions, implement gradient descent with fixed step sizes 0.1, 0.01 and 0.3 starting from the co-ordinate $(2,3)$, and any other 2 random points within their respective domains. Run 1000 iterations. For every function, every step size, and every initialization, plot the trajectory on top of the contour of the function (similar to Figure 1b). What can you say about the convergence of each function (in 1-2 lines, relate it to the theorems you have seen in class)?

3. *(10 marks)* Implement gradient descent with backtracking line search from the same starting points. The initial step size is 1, the backtrack ratio (β) is 0.5, the slope ratio (α) is 0.5, and the maximum number of backtracks is 10 (This is an upper limit on number of updates for $t := \beta t$). Run 1000 iterations and plot one trajectory on top of the contour for every function and every initialization. What can you say about the convergence of each function (in 1-2 lines, relate it to the theorems you have seen in class)?

4. *(10 marks)* Implement gradient descent from the same starting points with step size $= {}^1\!/_k$, where k is the current count of iteration. Run 1000 iterations and plot one trajectory on top of the contour for every function and every initialization. What can you say about the convergence of each function (in 1-2 lines, relate it to the theorems you have seen in class)?

5. *(4 marks)* Explain why for some problems large step sizes don't work. [Hint: check the eigenvalues of the Hessian.]