

# 6.867 Homework 1

Anonymous Authors

## I. IMPLEMENTING GRADIENT DESCENT

In any gradient descent algorithm, the main hyperparameters we have to tune are the initial point we start the gradient descent from, the step size, and the convergence criteria.

- An incorrect initial guess could lead to getting stuck at a local min without ever reaching the global minimum
- A step size that is too large can shoot past the minimum or go back and forth without ever reaching the critical point. Conversely, a step size that is too small can make the algorithm take far too long to converge.
- A convergence criteria that is too lax can result in a sub-optimal stopping point, while a convergence criteria that is too strict can result in the algorithm taking too long.

For each of the three parameters, we can see how varying the parameters changes the gradient descent for both of the provided functions (the Gaussian and the bowl).

Default parameters are:

- Starting Point: (0,0)
- Step Size: 0.01
- Convergence Criteria: Difference between consecutive objective function values is less than  $10^{-10}$ , max of 20,000 iterations

Varying Starting Point for Gaussian			
Starting Point	Distance from Critical Point	Number of Steps to Convergence	
1	6	87837	787
2	7	78	5415
3	545	778	7507
4	545	18744	7560
5	88	788	6344

## II. LINEAR BASIS FUNCTION REGRESSION

- Closed-form Solution with Polynomial Basis*
- Gradient Descent Solution with Polynomial Basis*
- Closed-form Solution with Cosine Function Basis*

## III. RIDGE REGRESSION

## IV. SPARSITY AND LASSO