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 minumum
- A step size that is too large can shoot past the minumum 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

II. Linear Basis Function Regression

A. Closed-form Solution with Polynomial Basis

Consider the basis function

$$\Phi_M(x) = [\phi_0(x), \ldots, \phi_M(x)],$$

where $\phi_k(x) = x^k$. Applying Φ_M to $\mathbf{X}$ gives the desired basis change, so we have the generalized linear model

$$\mathbf{Y} = \Phi_M(\mathbf{X}) \cdot \boldsymbol{\beta} + \boldsymbol{\epsilon},$$

which has the close-form solution

$$\hat{\boldsymbol{\beta}} = (\Phi_M(\mathbf{X})^T \Phi_M(\mathbf{X}))^{-1} \Phi_M(\mathbf{X})^T \mathbf{Y},$$

where $\hat{\boldsymbol{\beta}}$ is the maximum-likelihood estimator of the regression coefficients.

We ran regressions on the data using a few different degrees ($M = 0, 1, 3, 10$).[1] Below are plots of the resulting polynomial functions, compared to the given data and the true function:

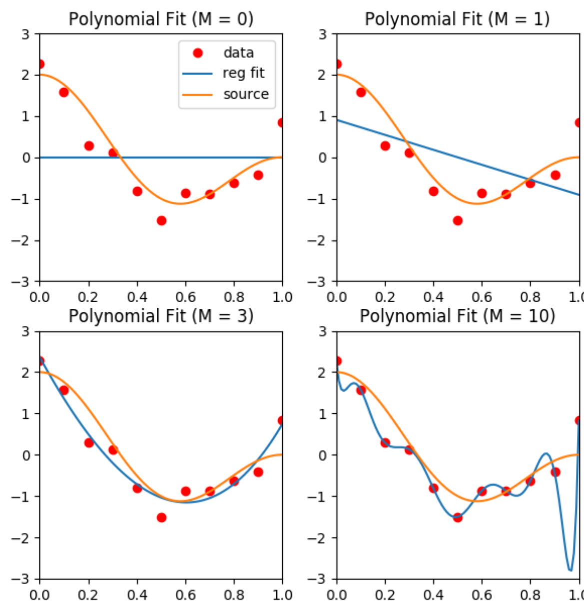

Fig. 1: M-Degree Polynomial Fit (M = 0, 1, 3, 10)

B. Gradient Descent Solution with Polynomial Basis

C. Closed-form Solution with Cosine Function Basis

III. Ridge Regression

IV. Sparsity and Lasso

V. Appendix

[1] See the Appendix for the numerical values of the weights.