STAT 102B: Homework 1

George Michailidis

Due electronically through **BruinLearn**
on April 14 at 5 pm

Problem 1:

Consider the function $f : \mathbb{R} \to \mathbb{R}$ with

$$f(x) = x^4 + 2x^2 + 1$$

Part (a):

Obtain the theoretical minimum of function $f(x)$. Show your work (recall to argue that it is actually a minimum).

Part (b):

Use the gradient descent algorithm **with constant step size** and **with backtracking line search** to calculate $\hat{x}_{\min}$.

1. For the constant step size version of gradient descent, discuss how you selected the step size used in your code.
2. For both versions of the gradient descent algorithm, plot the value of $f(x_k)$ as a function of k the number of iterations.
3. For the the gradient descent method with backtracking line search, plot the step size η_k selected at step k as a function of k. Comment on the result.

Problem 2: To understand the sensitivity of the gradient descent algorithm and its variants to the "shape" of the function, the two data sets provided (dataset1.csv, dataset2.csv) will be used.

They contain 100 observations for a response y and 20 predictors $x_j, j = 1, \cdots, 20$.

Part (a): Using the gradient descent code provided (both in `R` and `Python`) obtain the estimates of the regression coefficient, using both a constant step size and backtracking line search.

- Discuss how you selected the constant step size. Also, discuss which convergence criterion you used and the tolerance parameter used.
- Compare the results with those obtained from the `lm` command in `R` or from the class LinearRegression from the `sklearn.linear_model` in `Python`. Specifically, calculate $\|\hat{\beta}_{GD} - \hat{\beta}\|_2$, where $\hat{\beta}_{GD}$ is the estimate of the regression coefficient obtained from the gradient descent algorithm (both with constant step size and backtracking line search) and $\hat{\beta}$ obtained from the least squares solution implemented in `R` or `Python`.
- Plot the value of the objective function as a function of the number of iterations required.

Part (b): Implement the Polyak and Nesterov momentum methods and obtain the estimates of the regression coefficients, using both a constant step size and backtracking line search.

- Compare again the estimates obtained from the two momentum methods with the least-squares solution by calculating $\|\hat{\beta}_{GD} - \hat{\beta}\|_2$.
- Plot the value of the objective function as a function of the number of iterations required, and comment whether the momentum methods reduce the number of iterations requires to obtain the regression coeffcients (using the same tolerance).
- Comment on the results; namely, the difference in the accuracy of the solution and the number of iterations required by the momentum methods, versus the standard gradient descent algorithm.