
DATA130026.01 Optimization

Assignment 10

Due Time: at the beginning of the class, May 18, 2023

1. Consider the following logistic regression problem:

$$\min_{w \in \mathbb{R}^n, c \in \mathbb{R}} \frac{1}{m} \sum_{i=1}^m \log(1 + \exp(-b_i w^T a_i + c)) + \frac{1}{100m} \|w\|^2$$

where a_i, b_i are given data. Use the following MATLAB code to generate the data:

```
m = 500; n = 1000;
```

```
A = randn(n,m); b = sign(rand(m,1)-0.5);
```

Use your student ID as the random seed and the zero vector as starting point. Terminate your code when the Euclidean norm of the gradient is smaller than 10^{-4} or the maximum iteration number 3000 is reached. Implement gradient descent method to solve this problem and compare the convergence rate. You need to implement the line search gradient descent and constant step size gradient descent with at least two small fixed step size such that the algorithms can converge. Plot the results in one figure.

2. Consider the following logistic regression problem:

$$\min_{w \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m \log(1 + \exp(-b_i w^T a_i)) + \frac{1}{100m} \|w\|^2$$

where a_i, b_i are given. Test your algorithms on three real datasets from libsvm: 'a9a.test', 'CINA.test' and 'ijcnn1.test'. You may use the following MATLAB code to generate the data:

```
dataset = 'a9a.test';  
[b,A] = libsvmread(dataset);  
[m,n] = size(A);  
mu = 1e-2/m;
```

- (a) Use the zero vector as starting point. Terminate your code when the Euclidean norm of the gradient is smaller than 10^{-6} . Implement Newton's method and inexact Newton (Newton-CG) method (with three different inexact rules) to solve this problem and compare the convergence rate. The three different tolerance rules for inexact Newton methods are $\eta = \min(0.5, \|g\|)\|g\|$, $\eta = \min(0.5, \sqrt{\|g\|})\|g\|$, and $\eta = 0.5\|g\|$. You need to use Armijo rule line search for both methods. Plot the results (four lines) on gradient norm in one figure, and on function values $l(x_k) - l(x^*)$, for each dataset. ($l(x^*)$ can be obtained by running Newton's method for higher accuracy.)
- (b) Implement gradient descent with Armijo rule line search. Terminate your code with maximum iteration number 1000. Plot the results, together with two Newton's methods in the previous question, in one figure for each dataset.