MTH 410/510: HOMEWORK 4

PART I: Use the data set "LogisticData.txt" from d2l to implement the gradient decent method and the Newton method for the Logistic Regression. Find the coefficients w and b using the stopping criterion $\|\nabla L(x)\| < \epsilon$, where $\epsilon = 10^{-4}$. How many iterations are needed for each method?

PART II:

Problem 1. Consider the following optimization problem:

$$\begin{aligned}&\text{minimize } f(x_1, x_2, x_3) = x_1^2 + 2x_1x_2 + 3x_2^2 + 4x_1 + 5x_2 + 6x_3\\&\text{subject to } x_1 + 2x_2 = 3, 4x_1 + 5x_3 = 6.\end{aligned}$$

(a) Show that the function $f : I\!R^3 \to I\!R$ is convex.
(b) Find an absolute minimizer of the problem.
(Hint: Show that the Hessian of the function f is positive semidefinite and then use the method of Lagrange multipliers).

Problem 2. Consider the problem

$$\begin{aligned}&\text{minimize } \frac{1}{2}x^TQx - c^Tx + d\\&\text{subject to } Ax = b,\end{aligned}$$

where $Q^T = Q$ is an $n \times n$ positive definite matrix, A is an $m \times n$ matrix with rank $A = m$, $c \in I\!R^n$, and d is a constant. Derive a closed-form solution to the problem in terms of Q, c, d, A, b.

Problem 3. Find the absolute maximum of the problem

$$\begin{aligned}&\text{maximize } x + 2y\\&\text{subject to } x^2 + y^2 \le 1.\end{aligned}$$

(Hint: Convert the problem to a minimization problem and then apply the method of Lagrange multipliers).

Problem 4. Find the absolute minimum of the problem

$$\begin{aligned}&\text{mminimize } x_1^2 + x_2^2\\&\text{subject to } x_1 \ge 0, x_2 \ge 0, x_1 + x_2 \ge 5.\end{aligned}$$

(Hint: Convert the constraint to the right form and then apply the method of Lagrange multipliers).

Problem 5(bonus). Let $X_1, X_2, \ldots, X_n$ be independent random variables with $E(X_i) = \mu$ and Var $X_i = \sigma_i^2$. Find the coefficients a_i that minimize

$$\begin{aligned}&\text{Var } (\sum_{i=1}^n a_iX_i)\\&\text{subject to } E(\sum_{i=1}^n a_iX_i) = \mu.\end{aligned}$$

(Hint: Use Var $(aX) = a^2$Var X, Var $(X + Y)$ = Var (X) + Var (Y) for independent random variables X and Y.)

```
PART 1
Results with Gradient Descent ##############################
Elapsed time is 0.247191 seconds.
Coeffs. for w,b:
    1.5045
   -4.0774
Iterations: 3005
Prediction:
    0.0347
    0.0498
    0.0709
    0.1000
    0.1394
    0.1909
    0.1909
    0.2557
    0.3335
    0.4216
    0.5150
    0.6074
    0.6926
    0.7665
    0.8744
    0.9103
    0.9366
    0.9556
    0.9691
    0.9852
Magnitude of total errors: (Is this a useful thing to interpret?)
    1.6534

Results with Newton's Method ##############################
Elapsed time is 0.012428 seconds.
Coeffs. for w,b:
    1.5046
   -4.0777
Iterations: 6
Prediction:
    0.0347
    0.0498
    0.0709
    0.1000
    0.1393
    0.1908
    0.1908
    0.2557
    0.3335
    0.4216
    0.5150
```

```
    0.6074
    0.6926
    0.7665
    0.8744
    0.9103
    0.9366
    0.9556
    0.9691
    0.9852
Magnitude of total errors:
    1.6534
>>
```

```
% Jim Vargas
% MTH 410 Homework 4
clc, format compact

disp(newline+""+newline+"PART 1");
A=load("LogisticData.txt"); A=A';
y=A(:,2); m=size(A,1);
O=ones(m,1); X=A(:,1); X=[X,O]; % Format Data
epsilon=10^-4;
w0=[0,0]';


disp("Results with Gradient Descent ##############################");
tic
[w1,n1]=GradMethod(w0,X,m,y,epsilon);
toc
disp("Coeffs. for w,b:"); disp(w1);
disp("Iterations: "+n1);

disp("Prediction:");
t=-X*w1;
for i=1:m
    t(i)=1/(1+exp(t(i)));
end
disp(t);
disp("Magnitude of total errors: (Is this a useful thing to interpret?)");
disp(norm(y-t));


disp(newline+"Results with Newton's Method ##############################");
tic
[w2,n2]=NewtonMethod(w0,X,m,y,epsilon);
toc
disp("Coeffs. for w,b:"); disp(w2);
disp("Iterations: "+n2);

disp("Prediction:");
t=-X*w2;
for i=1:m
    t(i)=1/(1+exp(t(i)));
end
disp(t);
disp("Magnitude of total errors:");
disp(norm(y-t));

% figure
% plot(1:length(X),X,"."); title("X data");
```

Part 2 Jim Vargas MTH 410 HW4

Let it be assumed that I will be using the method of Lagrange multipliers for all of the following problems.

1) a)

Given $f(x) = x_1^2+2x_1x_2+3x_2^2+4x_1+5x_2+6x_3$ for $x \in \mathbb{R}^3$, the gradient and Hessian of f are, respectively,

$$\nabla f(x) = \begin{pmatrix} 2x_1 + 2x_2 + 4 \\ 2x_1 + 6x_2 + 5 \\ 6 \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 6 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It follows that $\det \nabla^2 f(x) = 0$ and $\operatorname{tr} \nabla^2 f(x) = 8$, and so f is convex.

1) b)

The Lagrangian function for the problem is

$$\mathscr{L}(x, \lambda) = f(x) - \lambda_1(x_1 + 2x_2 - 3) - \lambda_2(4x_1 + 5x_3 - 6)$$

with $\lambda = (\lambda_1, \lambda_2)$. Differentiating component-wise with respect to x and setting each to zero to find a unique minimizer x, in conjunction with the complimentary slackness conditions, we get the following equations:

$$\begin{aligned} 0 &= 2x_1 + 2x_2 - \lambda_1 - 4\lambda_2 + 4 \\ 0 &= 2x_1 + 6x_2 - 2\lambda_1 + 5 \\ 0 &= -5\lambda_2 + 6 \\ 0 &= \lambda_1(x_1 + 2x_2 - 3) \\ 0 &= \lambda_2(4x_1 + 5x_3 - 6). \end{aligned}$$

The third equation yields $\lambda_2 = \frac{6}{5}$. After substituting this and doing some manipulations, we get $x_1 = \frac{\lambda_1}{4} + \frac{37}{20}$ and $x_2 = \frac{\lambda_1}{4} - \frac{29}{20}$ and $x_3 = \frac{-\lambda_1}{5} - \frac{7}{24}$. Eventually more computation yields $\lambda_1 = \frac{27}{5}$ or $\lambda_1 = 0$, but $\lambda_1 = 0$ does not satisfy the constraints. Finally, the optimal solution is $x = \left(\frac{16}{5}, \frac{-1}{10}, \frac{-34}{24}\right)$ with $f(x) = \frac{1377}{100}$.

2)

The Lagrangian function for this problem can be written as

$$\mathscr{L}(x, \lambda) = \frac{1}{2}\left\langle Q^\top x, x\right\rangle - c^\top x + d - \langle \lambda A^\top, x\rangle + \langle \lambda, b\rangle.$$

Taking the gradient of $\mathscr{L}$ with respect to x and setting it to zero, we get

$$\begin{aligned} 0 &= Q^\top x - c^\top - \lambda A^\top \\ \Rightarrow x &= Q^{-1}(c^\top + \lambda A^\top). \end{aligned}$$

With the domain restriction, this means (at least, I think)

$$\begin{aligned} b &= A(Q^{-1}(c^\top + \lambda A^\top)) \\ \Rightarrow \lambda &= (AQ^{-1}A^\top)^{-1}(b - AQ^{-1}c^\top) \\ \Rightarrow x &= Q^{-1}(c^\top + [(AQ^{-1}A^\top)^{-1}(b - AQ^{-1}c^\top)]A^\top). \end{aligned}$$

3)

The Lagrangian function for this problem is

$$\mathscr{L}(x, y, \lambda) = x + 2y - \lambda(x^2 + y^2 - 1).$$

Differentiating with respect to x and y and setting to zero, along with the complimentary slackness condition, yields

$$\begin{aligned}
0 &= 1 - 2\lambda x \\
0 &= 2 - 2\lambda y \\
0 &= \lambda(x^2 + y^2 - 1).
\end{aligned}$$

With these, we get $x = \dfrac{1}{2\lambda}$ and $y = \dfrac{1}{\lambda}$ with $\lambda^2 = \dfrac{5}{4}$. Taking the positive root of λ to maximize $f(x, y) = x+2y$, the solution is $(x, y) = \left(\dfrac{1}{\sqrt{5}}, \dfrac{2}{\sqrt{5}}\right)$ with $f(x, y) = \sqrt{5}$.

4)

The Lagrangian function for this problem is

$$\mathscr{L}(x, \lambda) = x_1^2 + x_2^2 - \lambda_1(-x_1) - \lambda_2(-x_2) - \lambda_3(5 - x_1 - x_2)$$

for $x \in \mathbb{R}^2$, $\lambda = (\lambda_1, \lambda_2, \lambda_3)$. Taking the derivative component-wise and setting them to zero, after some computations we get $x_1 = \dfrac{-\lambda_1 - \lambda_3}{2}$ and $x_2 = \dfrac{-\lambda_2 - \lambda_3}{2}$ with the following conditions on λ_1, λ_2, λ_3:

$$\begin{aligned}
0 &= \lambda_1 \left(\frac{-\lambda_1 - \lambda_3}{2}\right) \\
0 &= \lambda_2 \left(\frac{-\lambda_2 - \lambda_3}{2}\right) \\
0 &= \lambda_3 \left(5 - \frac{-\lambda_1 - \lambda_3}{2} - \frac{-\lambda_2 - \lambda_3}{2}\right).
\end{aligned}$$

There are a few possible combinations of the lambdas which satisfy the above system alone, however, simple methods will show that only $\lambda_3 = -5$, $\lambda_1 = \lambda_2 = 0$ will satisfy the given boundary conditions. Therefore the solution is $x = \left(\dfrac{5}{2}, \dfrac{5}{2}\right)$ with $f(x) = \dfrac{25}{2}$.