Andrea Domuro Casanova
309354
Group 1: Tuesday Tutorial

EXERCISE SHEET 5

1. Newton Method

a) Newton method is an iterative algorithm for finding the points where a function $f(x)$ is egal to 0. That is, we want to minimize this function over some iterations.

Newton method is a second order method, it means that it uses the second derivate, ~~and~~

- Starting by a ~~method~~ univariate method, where we want to find the ~~0~~ zero.

$$x_{t+1} = x_t - \frac{f(x_t)}{f'(x_t)}$$

To minimize x_{t+1}, we ~~minim~~ iterate $x_t - \frac{f(x_t)}{f'(x_t)}$ until it converges. The purpose of this function is to find the 0, where the function crosses the x-axis.

Using this procedure ~~for minimizing~~ ~~we, now~~ now we want to minimize by finding where the first derivate is 0. (In the previous formula we wanted to find where the $f(x)$ is 0).

Basically, for ~~method~~ Newton's Method, is exactly the same but instead of using $f(x)$, we use $f'(x)$.

$$x_{t+1} = x_t - \frac{f'(x_t)}{f''(x_t)}.$$

b) $f_1(x) = x^3 - 2x - 5$

$\nabla f_1(x) = 3x^2 - 2\cancel{x}$

$\nabla^2 f_1(x) = 6x \cancel{-2}$

$f_2(x) = 3x^{1/3}$

$\nabla f_2(x) = \cancel{3} \cdot \frac{1}{\cancel{3}} x^{\frac{1}{3}-1} = x^{-\frac{2}{3}}$

$\nabla^2 f_2(x) = -\frac{2}{3} x^{-\frac{5}{3}}$

$$\Delta x^{k} = -(6x \cancel{-2})^{-1} \cdot (3x^2 - 2\cancel{x}) = \frac{-3x^2 + 2x}{6x - 2}$$

we use $\mu = 1$.

$$x^{(k+1)} = x^k + \mu \Delta x^k = x^k + \frac{-3x^{k2} + \cancel{2x^k}}{6x^k \cancel{-2}}$$

∘ Initial $x=8$ for $f_1(x)$:

$x^0 = 8$

$x1 = x^0 + \frac{-3 \cdot x^{0^2} + \cancel{2x^0}}{6x^0 \cancel{-2}} = 8 + \frac{-3 \cdot 8^2 + \cancel{2 \cdot 8}}{6 \cdot 8 \cancel{-2}} = 8 + \frac{-192 + \overset{2}{\cancel{16}}}{\underset{48}{\cancel{46}}} = 8 - 3.83 = \underline{\underline{\cancel{4.17}}}.$

$x^2 = 4.17 + \frac{-3 \cdot (4.17^2) + 2 \cdot 4.17}{6 \cdot 4.17 - 2} = 4.17 - \frac{43.82}{23.02} = 4.17 - 1.9 = \underline{\underline{2.27}}$

$x^3 = 2.27 + \frac{-3 \cdot 2.27^2 + 2 \times 2.27}{6 \times 2.27 - 2} = 2.27 - \frac{10.92}{11.62} = 2.27 - 0.94 = \underline{\underline{1.33}}$

$x^4 = 1.33 + \frac{-3 \cdot 1.33^2 + 2 \cdot 1.33}{6 \cdot 1.33 - 2} = 1.33 - \frac{2.65}{5.98} = 1.33 - 0.44 = \underline{\underline{0.89}}$

∘ Initial $x=-10$ for $f_1(x)$:

$x^0 = -10$

$x^1 = -10 + \frac{-3 \cdot (-10)^2 + 2 (-10)}{6 (-10) - 2} = -10 + \frac{320}{62} = -10 + 5.16 = \underline{\underline{-4.84}}$

$x^2 = -4.84 + \frac{-3 \cdot (-4.84)^2 + 2 \cdot (-4.84)}{6 (-4.84) - 2} = -4.84 + \frac{79.95}{31.04} = -4.84 + 2.57 = \underline{\underline{-2.27}}$

$x^3 = -2.27 + \frac{-3 \cdot (-2.27)^2 + 2 \cdot (-2.27)}{6 (-2.27) - 2} = -2.27 + \frac{19.99}{15.62} = -2.27 + 1.28 = \underline{\underline{-0.99}}$

$x^4 = -0.99 + \frac{-3 (-0.99)^2 + 2 \cdot (-0.99)}{6 (-0.99) - 2} = -0.99 + \frac{4.65}{7.94} = -0.99 + 0.58 = \underline{\underline{-0.41}}$

2.1

It seems that for both starting points, the updated points are getting close to 0.

$\nabla x^k = -(-\frac{2}{3} x^{-\frac{5}{3}})^{-1} \circ x^{-\frac{2}{3}} = -\frac{x^{-\frac{2}{3}}}{-\frac{2}{3} x^{-5/3}} = \frac{3}{2} x^{-\frac{2}{3} - (-\frac{5}{3})} = \frac{3}{2} x$

$x^{k+1} = x^k + 1 \cdot \frac{3}{2} x^k$

∘Initial value $x = -0.5$ for $f_2(x)$:

$x^0 = -0.5$

$x^1 = -0.5 + \underbrace{\frac{3}{2}}_{1.5}(-0.5) = -0.5 - 0.75 = -1.25$

$x^2 = -1.25 + 1.5 \cdot (-1.25) = -3.125$

$x^3 = -3.125 + 1.5 \times (-3.125) = -7.81$

$x^4 = -7.81 + 1.5 \times (-7.81) = -19.525$

∘Initial value ~~for~~ $x = 1$ for $f_2(x)$:

$x^0 = 1$

$x^1 = 1 + \frac{3}{2} \cdot 1 = 2.5$

$x^2 = 2.5 + 1.5 \cdot 2.5 = 6.25$

$x^3 = 6.25 + 1.5 \times 6.25 = 15.62$

$x^4 = 15.62 + 1.5 \times 15.62 = 39.05$

For $f_2(x)$ seems that if the initial value is negative, then for each iteration, the updates values will get smaller. But when the starting value is positive, then we increment the updated value for each iteration.

c) Newton's method fails in the cases where the derivate is 0. When the derivate is very close to 0, the tangent line is nearly horizontal and then ~~the~~ it could overshoot.

3.1

2. Newton Method for ML problems.

a) $\mathcal{L}(X,\beta,Y) = \sum_{i=1}^{m} (x_i\beta - y_i)^2$

$$\frac{\delta \mathcal{L}(X,\beta,Y)}{\delta\beta} = -2X^T(y-X\beta)$$

$$\nabla^2 \mathcal{L}(X,\beta,Y) = 2X^T X$$

We could use Newton minimization algorithm but the algorithms just end up being the Least Squares.

$$\frac{\nabla f(x)}{\nabla^2 f(x)} = \frac{-\cancel{2}\cancel{X^T}(y-X\beta)}{\cancel{2}\cancel{X^T}X} = -(X^T)^{-1}(y-X\beta)$$

Since it is the least Square, we don't really need the newton method because we can just solve this equation system. So it wouldn't make sense.

b) $\mathcal{L}(X,\beta,Y) = -\sum_{i=1}^{m} y_i \log(\sigma(x_i\beta)) + (1-y_i)\log(1-\sigma(x_i\beta))$

$\nabla \mathcal{L}(X,\beta,Y) = -\sum x_i(y_i - \sigma(x_i\beta))$

dot product.

$H = \nabla^2 \mathcal{L}(X,\beta,Y) = -X^T W X$, where $W = \operatorname{diag}(\underbrace{f_\theta(x)}_{f_\theta(x) = \frac{e^{x_n^T\theta}}{1+e^{x_n^T\theta}}} \odot (1-f_\theta(x)))$

4.1

W is a diagonal matrix, where each value of the diagonal is $f_\theta(x) \odot \cancel{f_\theta(x)}) (1-f_\theta(x))$

For the ~~log~~ loss function of logistic regression, we want to ~~maximize~~ minimize the function, We want to solve a linear combination by using a smooth function and with a range from 0 to 1. From taking the second deriviate (Hessian), we make sure that we will find a global ~~minimum~~. optimal solution. And also, ~~with~~ the hessian matrix updates the direction and the Newton algorithm usually converges quicker.

Index der Kommentare

2.1 here there is a slight miscalculation

3.1 also when the function is oscilating(like sin wave for example)

4.1 here you should show your workout on reaching the hessian