$\frac{90}{90} = \frac{100}{100}$

CS 4365/CS 5354 Data Processing Under Security and Privacy Summer 2016, Test 1

10|10

Name: [redacted]

1. Similarly to how we used Newton's method to design algorithms for computing square root and cubic root, design an algorithm for computing the logarithm x = ln(a) as a solution to the equation

$$e^x = a.$$

Suppose we know the approximate solution $x^{(k)}$ at step K, then

$$x^{(k+1)} = x^{(k)} + \Delta x$$

According to Newton's method,

$$\Delta x = -\frac{f(x^{(k)})}{f'(x^{(k)})}$$

$$f(x) = e^x - a = 0$$

$$f'(x) = e^x$$

$$\therefore \Delta x = -\frac{e^x - a}{e^x} = \frac{a - e^x}{e^x} = \left(\frac{a}{e^x} - 1\right) \quad \text{— (i)}$$

① Approximate a value, for example, $x^{(0)} = 0$.

② compute Δx where $x = x^{(k)}$ [At first time $k=0$] using equation (i)

③ $x^{(k+1)} = x^{(k)} + \Delta x$

④ Repeat the step (2) to (3) until the approximation of x is ~~close~~ very close to solution meaning the error is less than a predefined tolerance.

$\frac{10}{10}$

2. Use the algorithm for computing 1/b that we had in class (and that is implemented in the computers) to perform the few first steps of computing the ratio 1 / 1.1.

Let $\frac{1}{b} = x$ ~~such that~~ $x \cdot b = 1$.

$$F(x) = x \cdot b - 1 = 0$$

$$F'(x) = b$$

using Newton's method,

$$\Delta = -\frac{F(x)}{F'(x)} = -\frac{x \cdot b - 1}{b} = \frac{1 - x \cdot b}{b}$$

$$= (1 - x \cdot b) \cdot \frac{1}{b}$$

In every step, x is the approximate value for $(\frac{1}{b})$. Therefore, we can write,

$$\Delta x = (1 - x \cdot b) \cdot x$$

So at step $k+1$,

$$x^{(k+1)} = x^{(k)} + (1 - x^{(k)} \cdot b) \cdot x^{(k)}$$

$$= x^{(k)} (2 - x^{(k)} \cdot b) \longrightarrow (i)$$

In the problem, $b = 1.1$

Let $x^{(0)} = 1$,

Iteration 1, $x^{(1)} = 0.9$

Iteration 2, $x^{(2)} = 0.909$

Iteration 3, $x^{(3)} = 0.9090909$

$x^{(3)}$ is very close to solution, so we can stop here.

$$x = 0.9090909.$$

3-6. 40/40

3. Use numerical differentiation to compute the derivative of the function $x^2 - x$ when $x = 1$. 10/10

4. Use linearization technique and your estimate for the derivative to estimate the range of this function when x is in the interval [0.9, 1.1]. 10/10

5. Use naive interval computations to estimate the same range. 10/10

6. Use mean value form to estimate the same range. 10/10

3. By definition of differentiation,

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \longrightarrow (i)$$

for some small value of $h \ll 1$, we can write,

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

Here, $f(x) = x^2 - x$

we want to find $f'(x)$ at point $x = 1$.

Let $h = 0.001$.

$$f'(x) = \frac{f(1.001) - f(1)}{0.001}$$

$$= 1.001$$

if $h = 0.00001$,

$$f'(x) = \frac{f(1.00001) - f(1)}{0.0001} = 1.00001$$

Analytical solution: $f'(1) = 1$, which is very close to numerical differentiation.

4. Here given that.

$$y = f(x) = x^2 - x \text{ where}$$

$$x \in [0.9, 1.1]$$

$$\hat{x} = \frac{\underline{x} + \overline{x}}{2} = \frac{0.9 + 1.1}{2} = 1$$

$$\Delta_x = \frac{\overline{x} - \underline{x}}{2} = \frac{1.1 - 0.9}{2} = 0.1$$

$$\hat{y} = f(\hat{x}) = 1^2 - 1 = 0$$

We know,

$$\Delta = \left|\frac{\partial f}{\partial x}\right| \Delta_x$$

$$= |(2x-1)| \cdot \Delta_x$$

$$= |(2 \cdot 1 - 1)| \cdot 0.1$$

$$= 1 * 0.1$$

$$= 0.1.$$

$\therefore$ The range of y $= [0 - 0.1, 0 + 0.1]$

$$= [-0.1, 0.1]$$

5.

$$f(x) = x^2 - x \quad \text{where } x \in [0.9, 1.1]$$

$$= [0.9, 1.1]^2 - [0.9, 1.1]$$

$$= [0.81, 1.21] - [0.9, 1.1]$$

$$= [-0.29, 0.31]$$

6. $f(x) = x^2 - x$ where $x \in [0.9, 1.1]$

$\Delta x = 0.1$, $\check{x} = 1$

$$\frac{\partial f}{\partial x} = 2x - 1$$

$$\check{y} = f(\check{x}) = 1^2 - 1 = 0$$

According to mean-value form

$$f(x) \in \check{y} + \left[\frac{\partial f}{\partial x}\right] \cdot [-\Delta x, \Delta x]$$

$$= 0 + [2x - 1][-\Delta x, \Delta x]$$

$$2x - 1 = 2[0.9, 1.1] - 1$$

$$= [1.8, 2.2] - 1$$

$$= [0.8, 1.2]$$

$$\therefore f(x) \in [0.8, 1.2][-0.1, 0.1]$$

$$= [-0.12, 0.12]$$

12/10

7. Use Newton's method to solve the following system of non-linear equations:

$$x_1 * x_2 = 3, x_1 + x_2 = 4.$$

Start with the first approximation $x_1 = 1$ and $x_2 = 2$. One iteration is good enough.

Given:

$$f_1(x_1, x_2) = x_1 * x_2 - 3 = 0$$

$$f_2(x_1, x_2) = x_1 + x_2 - 4 = 0 \quad \text{and } x_1^{(0)} = 1 \text{ and } x_2^{(0)} = 2$$

According to Newton's method for systems of non-linear equation, we can write.

$$\frac{\partial f_1}{\partial x_1} \cdot \Delta x_1 + \frac{\partial f_1}{\partial x_2} \Delta x_2 = -f_1(x_1^{(0)}, x_2^{(0)})$$

$$\frac{\partial f_2}{\partial x_1} \Delta x_1 + \frac{\partial f_2}{\partial x_2} \Delta x_2 = -f_2(x_1^{(0)}, x_2^{(0)})$$

$$\frac{\partial f_1}{\partial x_1} = x_2 \qquad \frac{\partial f_2}{\partial x_1} = 1$$

$$\frac{\partial f_1}{\partial x_2} = x_1 \qquad \frac{\partial f_2}{\partial x_2} = 1.$$

So, $f_1(x_1^{(0)}, x_2^{(0)}) = 1 * 2 - 3 = -1$

$f_2(x_1^{(0)}, x_2^{(0)}) = 1 + 2 - 4 = -1$

$$2\Delta x_1 + \Delta x_2 = 1$$

$$\Delta x_1 + \Delta x_2 = 1$$

$\therefore \Delta x_1 = 0$

and $\Delta x_2 = 1.$

$\therefore x_1^{(1)} = 1 + 0 = 1$

$x_2^{(1)} = 2 + 1 = 3.$

Solution, $x_1 = 1$ and $x_2 = 3.$

8. Find the point closest the origin on the line $x_1 - x_2 = 1$. In other words, find the values x_1 and x_2 for which the sum $(x_1)^2 + (x_2)^2$ attains the smallest possible value under the constraint $x_1 - x_2 = 1$.

10/10

$$f(x) = (x_1)^2 + (x_2)^2$$

$$g(x) = x_1 - x_2 - 1 = 0$$

using lagrange multiplier,

$$L(x) = (x_1)^2 + (x_2)^2 + \lambda(x_1 - x_2 - 1)$$

$$\frac{\partial L}{\partial x_1} = 2x_1 + \lambda = 0$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - \lambda = 0$$

$$\therefore x_1 = -\frac{\lambda}{2}$$

and $x_2 = \frac{\lambda}{2} = -x_1$

from $g(x) = 0$, putting the value of x_1 and x_2,

$$2x_1 = 1$$

$$x_1 = \frac{1}{2}$$

and $\cancel{-2}x_2 = 1$

$$\therefore x_2 = -\frac{1}{2}$$

$$(x_1, x_2) = \left(\frac{1}{2}, -\frac{1}{2}\right)$$

10/10

9. Explain what is k-anonymity, and why it is important. If k increases, will we get more or less privacy protection? Explain your answer.

Let a desired statistical characteristics for a dataset,

$$C(x_1^{(1)}, \ldots, x_n^{(1)}, \ldots x_1^{(N)}, \ldots x_n^{(N)}).$$

To protect privacy of the dataset, we can introduce interval instead of exact values. Assuming that the dataset distribute over a hyperspace, we can make cells for every characteristic to introduce interval.

K-anonymity is the condition where in every cell for different characteristics, there should be at least K different people after all queries.

This is the primary important concept in data privacy. Assume, from a people record dataset, we erase name of people. Still it's not private because if we know the scope/all people of the dataset, and allows to query for n-1 people, we can easily identify any private information of a particular person. K-anonymit solves this problem.

If K-increases, we will get more privacy because for every query we will find at least K-people, which is difficult to identify ~~the~~ a particular person.