OM Homework Set 5

1.1

$\boxed{1\cdot 1}$ Jensen's inequality

let $f(x)$ be convex in interval I

Then for any $\{x_i \mid i \in \{1 \text{ to } t\}\ \&\ x_i \in I\}$

$$f\left(\frac{\sum_{1}^{t} x_i}{|x_i|}\right) \leq \text{avg of } \{f(x_i)\}$$

and if f were concave, the inequality is reversed i.e

$$f\left(\frac{\sum_{1}^{t} x_i}{|x_i|}\right) \geq \text{avg of } \{f(x_i)\}$$

to prove : $\frac{1}{1-x} + \frac{1}{x} + \frac{1}{x+1} > \frac{3}{x}$

~~for~~ $f(x) = 1/x$ which is convex in interval $(0,\infty)$

let the set $X = \{(x-1), x, (x+1)\}$

$$\text{Avg of } \{f(x)\} = 1/3(f(x-1) + f(x) + f(x+1))$$

$$= \frac{1}{3}\left[\frac{1}{x-1} + \frac{1}{x} + \frac{1}{x+1}\right] \quad -(1)$$

and $f(\text{avg of } f\{x_i\}) = f\left(\frac{1}{3}(x-1+x+x+1)\right)$

$$= f(x)$$

$$= 1/x, \quad — \quad (2)$$

from 1 & 2.

$$1/3\,(1/x-1 + 1/x + 1/x+1) > 1/x$$

$$1/x-1 + 1/x + 1/x+1 > 3/x.$$

Scanned with CamScanner

→ given series $s = 1 + 1/2 + 1/3 + \dots + 1/n + \dots$

s can be written as

$$s = \left(\frac{1}{2-1} + \frac{1}{2} + \frac{1}{2+1}\right) + 1/4 + \left(\frac{1}{6-1} + 1/6 + \frac{1}{5+1}\right) + \frac{1}{8} + \left(\frac{1}{10-1} + \frac{1}{10} + \frac{1}{10+1}\right) + 1/12 + \dots$$

using term for number in for num in we get

$$s > (3/2 + 3/6 + 3/10 \dots) + (1/4 + 1/8 + 1/12 \dots)$$

$$s > 3/2\,(1 + 1/3 + 1/5 + 1/7 + \dots) + \tfrac{1}{4}\underbrace{(1 + 1/2 + 1/3 + 1/4 \dots)}_{s}$$

$$s > 3/2\,(1 + 1/3 + 1/5 + 1/7 + \dots) + s/4$$

$$s/4 > 1/2\,(1 + 1/3 + 1/5 \dots) \quad \text{--- (1)}$$

Now let $s_1 = 1 + (1/3) + (1/5 + 1/7) + (1/9 + 1/11 + 1/13 + 1/15) \dots$

$$= 1 + 1/4 + 1/4 + \dots$$

$$= 1 + \sum_{n=1}^{\infty} n/4$$

as $n \to \infty$ $\quad \sum n/4 \to \infty$

$s_1 \to \infty$ and hence s_1 is divergent

Now from (1)

$$s > 4/2\, s_1 \quad \text{or} \quad s > 2 s_1$$

Since s_1 is divergent which implies s is also divergent

∴ the given series does not converge for real number.

1.2

1.2 There chords lemma.

let $f: I \to R$ then f is convex on I if and only if.
and only if for any pts $a, b, c \in I$ with $a<b<c$
we have $\frac{f(b)-f(a)}{b-a} \le \frac{f(c)-f(a)}{c-a} \le \frac{f(c)-f(b)}{c-b}$

Proof: Given that f is convex on I & $a<b<c$
we can write $b = \lambda + (1-\lambda)c$.
where $\lambda = \frac{c-b}{c-a} \in [0,1]$

since f is convex.

$$f(b) \le \frac{(c-b)f(a)}{c-a} + \frac{(b-a)}{c-a} f(c) \quad \text{——} (1)$$

$$f(b)-f(a) \le \frac{b-a}{c-a}f(a) + \frac{b-a}{c-a} f(c).$$

$$f(b)-f(a) \le \frac{b-a}{b-c}(f(c)-f(a))$$

$$\frac{f(b)-f(a)}{b-a} \le \frac{f(c)-f(a)}{c-a} \quad \text{——} (2).$$

hence prove 1 inequality

from eq (1) we have.

$$f(b) < \lambda f(a) + (1-\lambda) f(c).$$

substitute value of λ

$$f(b) < \frac{c-b}{c-a}f(a) + \left(1-\frac{c-b}{c-a}\right) f(c)$$

$$\frac{c-b}{c-a}[f(c)-f(a)] \le f(c)-f(b)$$

$$\Rightarrow \frac{f(c)-f(a)}{c-a} \le \frac{f(c)-f(b)}{c-b} \quad \text{——} (3)$$

from 2 & 3.

$$\frac{f(b)-f(a)}{b-a} \le \frac{f(c)-f(a)}{c-b} \le \frac{f(c)-f(b)}{c-b}$$

Scanned with CamScanner

1.9 Given x, y, z are positive real numbers with $x+y+z = 1$

To show: $(1+1/x)(1+1/y)(1+1/z) \geq 64$

Sol^n: let $f(x) = \log(1+1/x)$ which is convex in $(0,\infty)$

let's consider the average of fun $f(x), f(y), f(z)$

$$\frac{1}{3}(f(x)+f(y)+f(z))$$

$$= \frac{\log(1+1/x) + \log(1+1/y) + \log(1+1/z)}{3}$$

$$= \frac{\log((1+1/x)(1+1/y)(1+1/z))}{3} \quad \text{--- (1)}$$

Now let,

$$f\left(\frac{x+y+z}{3}\right) = \log(1+3/(x+y+z)) = \log(4) \quad \text{--- (2)}$$

Now from jensen's equality:

$$\frac{1}{3}(f(x)+f(y)+f(z)) \geq f((x+y+z)/3)$$

from (1) & 2

$$\frac{\log[(1+1/x)(1+1/y)(1+1/z)]}{3} \geq \log_2 4.$$

Solving logarithm

$$\frac{1}{3}(1+1/x)(1+1/y)(1+1/z) \geq 64$$

1.4 Conduct 2 iterations of secant method

$x^2 - 2x - 5 = 0$ where $x_0 = 2$

$x_1 = 3$

soln: The equation for calculating root of a secant line is given as

$$x_2 = x_1 - \frac{(x_1 - x_0)}{f(x_1) - f(x_0)} f(x_1)$$

here $f(x) = x^3 - 2x - 5$

$f(x_0) = f(2) = -1$

$f(x_1) = f(3) = 16$

<u>iteration 1</u>

$$x_2 = 3 - \frac{(3-2)}{16-(-1)} 16$$

$$= 3 - \frac{16}{17} = \frac{51-16}{17} = \frac{35}{17}$$

$$= 2.0588$$

<u>iteration 2</u>

$$x_3 = x_2 - \frac{x_2 - x_1}{f(x_2) - f(x_1)} f(x_2)$$

$$= \frac{35}{17} - \frac{\left(\frac{35}{17} - 3\right)}{f(35/17) - 16} \cdot f(35/17)$$

Scanned with CamScanner

$$3 - 16 \cdot \frac{(2.0588 - 3)}{-3.3908 - 16}$$

$$x_3 = 2.0813$$

$$f(x_3) = f(2.0813) = -0.1472$$

Scanned with CamScanner

1.5 conduct 2 iterations of bisection method

$xe^x = 1$ where $x \in [0, 1]$

soln given

$x.e^x - 1 = 0$

iteration 1)

here $f(0) = 0e^0 - 1 = 0 - 1 < 0$

& $f(1) = e^1 - 1 = e - 1 > 0$

sign change form $0^{-ve} \to 1^{+ve}$

root $\in [0, 1]$

$x_0 = \frac{0+1}{2} = 0.5$

$f(x_0) = f(0.5) = 0.5 e^{0.5} - 1$

$= \frac{1}{2}\sqrt{e} - 1$

$= \left(\frac{\sqrt{e} - 2}{2}\right) = -0.175720$

iteration 2

$f(0.5) = -0.1757 < 0$

$= \frac{1}{2}\sqrt{e} - 1$

$= \frac{\sqrt{e} - 2}{2} = -0.175720$

Scanned with CamScanner

iteration 2

$f(0.5) = -0.1157 < 0$

$f(1) = e-1 > 0$

sign from -ve to +ve

$x_1 = \left(\frac{0.5+1}{2}\right) = 0.75$

$f(x_1) = (0.75)\, e^{0.75} - 1$

$= 3/4 \sqrt[4]{e^3} - 1$

$= 1.587 - 1$

$= 0.587$

$\therefore$

$$\boxed{\begin{aligned} f(x_0) &= -0.1157 \\ f(x) &= 0.587 \end{aligned}}$$

Scanned with CamScanner

~~Given: $x^2-2x-5=0$~~ 1.6

~~$x_1=2$, $x_2=3$~~

1.6 Find a real root of equation of $x^2-y^2=3$ & $x^2+y^2=13$ by doing 2 iterations of newton method

$x_0=y_0=\sqrt{6.5}$

soln Given, $x_0=y_0=\sqrt{6.5}=2.54$

$x^2-y^2-3=0$

$f_1(x,y) = x^2-y^2-3.$

$x^2+y^2-13=0$

$f(x,y) \quad x^2+y^2-13$

Jaccrobian $J=\begin{bmatrix}\frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y}\\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y}\end{bmatrix}$

$\Rightarrow \begin{bmatrix}2x & -2y\\ 2x & 2y\end{bmatrix}$

$f\begin{pmatrix}x_n\\ y_n\end{pmatrix} = \begin{bmatrix}f(x_n\ y_n)\\ f_2(x_n\ y_n)\end{bmatrix}$

1st iteration

$\begin{bmatrix}x_{n+1}\\ y_{n+1}\end{bmatrix} = \begin{bmatrix}x_n\\ y_n\end{bmatrix} - J^{+} f\begin{bmatrix}x_n\\ y_n\end{bmatrix}$

$=\begin{bmatrix}2.5495\\ 2.5495\end{bmatrix} - \begin{bmatrix}5.09902 & -5.09902\\ 5.09902 & 5.009902\end{bmatrix} * f\begin{bmatrix}2.5495\\ 2.5498\end{bmatrix}$

~~iteration 2) here n=1 $x_1 = 2.9436$ $x_2 = 2.2553$~~

~~$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - J^{-1} f \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$~~

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.8436 \\ 2.2553 \end{bmatrix}$$

iteration 2

$n=1$ $x_1 = 2.8436$ $x_2 = 2.2553$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - J^{+} f \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

$$= \begin{bmatrix} 2.8436 \\ 2.2553 \end{bmatrix} - \begin{bmatrix} 5.68737 & -4.5106 \\ 5.66737 & 4.5106 \end{bmatrix} \begin{bmatrix} 2.6436 \\ 2.2553 \end{bmatrix}$$

$$= \begin{bmatrix} 2.8247 \\ 2.23615 \end{bmatrix}$$

hence final answer. after 2 iteration

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.82847 \\ 2.23615 \end{bmatrix}$$

1.7 2 iteration.

$2x - \cos x - 3 = 0$

take $g(x) = \frac{\cos x + 3}{2}$ $\qquad x_0 = \pi/3$

sol^n given:-

$g(x) = \frac{\cos x + 3}{2}$ $\qquad x_0 = \pi/2$

Now according to fixed point method

$x_{n+1} = g(x_n)$

iteration 1 :

$x_0 = \pi/2$ $\qquad$ ~~$f(x_0) = \pi - 3 = 0.14159$~~

$x_1 = g(x_0)$

$= \frac{\cos(\pi/2) + 3}{2}$

$x_1 = 1.5$ $\qquad \boxed{f(x_1) = 0.3472}$

Iteration 2 :

$x_1 = 1.5$

$x_2 = g(1.5)$

$= \frac{\cos(1.5) + 3}{2} = \frac{0.0707 + 3}{2}$

$x_2 = 1.53537$, $f(x_2) = -0.0173$

Scanned with CamScanner

1.8

Pseudo code for gradient ascent

$x \rightarrow$ data of m samples, n features

$y \rightarrow$ output

$\alpha \rightarrow$ learning rate

$W \rightarrow$ learn values

$W_{n+1} = W_n + \alpha$

updated steps ~~give~~ gradient as shown above

Grad_desc (X, Y, R, num_items):

m, n = shape(X)

α ~~alpha~~ = R

W := n zeros.

for i = 1 to num_items:

begin

$\hat{y} = X * W$

error = $y - \hat{y}$

$W = W_0 + \alpha *$ error $* X$

End

Max $[6 - (x_1^2 (x_1^2 - 16) + x_2^2 (x_2^2 - 9))$

at $W_0 = (0,0)$, the gradient does not update & the max value of f^n stays at 6

- if the starting point is changed to [0.2, 0.1], the max value of the f^n reache is 90.25

1.9 Batch gradient descent: In this entire traning set is used to perform one iteration of gradient descent the average of the gradients of all the training example is taken and this mean gradient is used to update parameter

Advantage of gradient descent:

1. less ~~os~~ oscillation and noisy steps taken towards the global minima of the loss function due to updating parameter by computing the average of all the training sample rather than the value of single sample

2. it can benefit from the vectorization which increases the speed of processing all training samples together

3. it processes a more stable gradient descent convergence and stable of ~~processing all~~ error gradient that stochastic gradescent

4. mini batch GD: This uses a small subset of traning data to compute the gradient.

stochastic GD: A single training sample is used to find gradient and update parameter

5. Both GD is used to because it is good for conuex or relatively smooth error manifolds as it moves somewhat directly towards an optimum soln

But using SGD or minibatch GD the optimization path taken is erratic and give inaccurate but with advantage.

Scanned with CamScanner

1.10 Convergence proof of gradient descent

→ suppose the fn $f: R^n \to R$ is convex and differentiable, and that it gradient is lipschitz continuous with constant $L>0$ i.e $||\nabla f(x) - \nabla f(y)|| \leq L ||x-y||_2$ for any x, y.

then if grad. descent is run for k iteration with a fixed step $t \leq 1/2$ it will yeild a sol^n satisfies

$$f(x^k) - f(x^*) \leq \frac{||x^k - x^*||^2}{2t k} \quad \text{——— (1)}$$

where $f(x^*)$ is optimal value, this means GD is to converge with rate $1/K$

proof: Since ∇f is lipschitz continuous with const L

$$\Rightarrow \nabla^2 f(x) \leq LI$$

or $\nabla^2 f$ LI is negative semidefinite matrix

using quadratic expansion

$$\nabla f(y) \leq f(x) + \nabla f(x)^T (y-x) + \tfrac{1}{2} \nabla^2 f(x) ||y-x||^2$$
$$\leq f(x) + \nabla f(x)^T (y-x) + \tfrac{1}{2} L ||y-x||_2^2.$$

by putting grad. des update in this

$y = x^+ = x - t\nabla f(x)$

$$f(x^+) \leq f(x) + \nabla f(x)^T (x^+ - x) + \tfrac{1}{2} L ||x^+ - x||^2$$
$$= f(x) + \nabla f(x)^T (x - t\nabla f(x) - x)^2 + \tfrac{1}{2} L ||x - t\nabla f(x) - x||^2$$
$$= f(x) + \nabla f(x)^T + \nabla f(x) + \tfrac{1}{2} L ||t \nabla f(x)||^2$$
$$= f(x) - t||\nabla f(x)||_2^2 + \tfrac{1}{2} L t^2 ||\nabla f(x)||_2^2$$
$$= f(x) - (1 - \tfrac{1}{2}) L t + ||\nabla f(x)||_2^2$$
$$= f(x) - (1 - \tfrac{1}{2} L t)^2 + ||\nabla f(x)||_2^2 \quad \text{——— (2)}$$

using $t \leq 1/L$ we know that

$$t(1 - \tfrac{1}{2} L t) = \tfrac{1}{2} L t - 1 \leq \tfrac{1}{2} L (1/L) t = \tfrac{1}{2} - 1 = -\tfrac{1}{2}$$

in eq (2)

$$f(x^+) \leq f(x) - \tfrac{1}{2} t ||\nabla f(x)||_2^2 \quad \text{— (3)}$$

since $\frac{1}{2}t\|\nabla f(x)\|$ will always be +ve unless $\nabla f(x)=0$ this inequality implies that objective fn. value strictly decreases with each iteration of grad desc until it reaches the optimum $f(x)=f(x^*)$. this holds only if t selected such as $t\leq 1/L$

We can bound $f(x^+)$, the objective value at next iteration in terms of $f(x^*)$ the optimal objective value, since f is convex

$$f(x^*) \geq f(x) + \nabla f(x)^T(x^*-x)$$

$$f(x) \leq f(x^*) + \nabla f(x)^T(x-x^*)$$

putting this in (3)

$$f(x^+) \leq f(x^*) + \nabla f(x)^T(x-x^*) - t/2\,\|\nabla f(x)\|_2^2$$

$$f(x^+) - f(x^*) \leq \frac{1}{2t}\left(2t\,\nabla f(x)^T(x-x^*) - t^2\|\nabla f(x)\|_2^2\right)$$

$$f(x^+) - f(x) \leq 1/2t\left(\|x-x^*\|_2^2 - \|x-t\nabla f(x)-x^*\|_2^2\right) \quad (4)$$

by definition $x^+ = x - t\nabla f(x)$ putting this in (4)

$$f(x^+) - f(x^*) \leq 1/2t\left(\|x-x^*\|_2^2 - \|x^+-x^*\|_2^2\right) \quad (5)$$

this inequality holds for x^+ in every iteration of G.D.

$$\sum_{i=1}^{k} f(x^{(i)}) - f(x^*) \leq \sum_{i=1}^{k}\frac{1}{2t}\left(\|x^{i-1}-x^*\|_2^2 - \|x^i-x^*\|_2^2\right)$$

$$= 1/2t\left(\|x^0-x^*\|_2^2 - \|x^k-x^*\|_2^2\right)$$

$$\leq 1/2t\left(\|x^0-x^*\|_2^2\right)$$

Now f is decreasing on every iteration we can conclude that $f(x^{(k)}) - f(x^*) \leq 1/k\sum_{i=1}^{k} f(x^i) - f(x^*)$

$$\leq \left\|\frac{x^{(0)}-x^*}{2tk}\right\|^2 \quad \text{(stability from (})$$

hence proves (1) the convergence of gradient descent

Scanned with CamScanner