Theory Questions

Date....../....../......

ISHMEET KAUR
2015042

4. $$p(x;\theta) = \begin{cases} \frac{1}{\pi\theta^2} & ; \quad \|x\| \le \theta \\ 0 & ; \quad \text{otherwise} \end{cases}$$

The likelihood function will be:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\pi\theta^2} = \left(\frac{1}{\pi\theta^2}\right)^n$$

Log likelihood function:

$$l(\theta) = \log\left(\left(\frac{1}{\pi\theta^2}\right)^n\right)$$

$$= -n\pi \log\theta^2$$

$$= -2\pi n \log\theta$$

we need to maximise this function.
$\log\theta$ is an increasing function. So, we see, as θ increases, log-likelihood decreases.
Hence, θ should be as small as possible.
But, given that

$$\theta \geq \|x\| \quad \text{i.e.}$$ (else likelihood will be 0)

$$\theta \geq \max_{i=1}^{n} \|x_i\|$$

So, maximum likelihood estimation of θ is:

$$\boxed{\hat{\theta} = \max_{i=1}^{n} \|x_i\|}$$

Date....../....../......

1. Gradient search solution – $O(n)$ time
Find minima (Hessian method) — $O(n^3)$ time
several iterations run in gradient descent, wheras in Newton Descent, time = $O(n^3)$. but it is faster for a smaller set of data. For larger datasets, gradient descent still runs in $O(n)$ time. And there is no space issue too.

2. In function approximation it is very conviniently assumed that the function necessarily overfits/ underfits the data and that there is large error in the test data. But, in machine learning, we don't have overfitting/ underfitting to that extent as we try to keep our parameters such that there is no overfitting or underfitting. Also, with machine learning, we aim at figuring out the patterns & learning upon them, so, the whole ideas of the two things is very different.