Macline Learning HW - 1

Theory.

SHREYA SHARMA
2015096

Q4. we are given $p(x;\theta) = \begin{cases} \frac{1}{\pi\theta^2} ; & \|x\| \le \theta \\ 0 ; & \text{otherwise} \end{cases}$

$$\|x\| = \sqrt{x_1^2 + x_2^2}$$

$L(\theta)$: likelihood function :

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

$l(\theta)$: log likelihood function

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

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

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

$$= -n\log\theta^2 - \text{const}$$

$$= -2n\log\theta^2 + \text{const}$$

$\max(l(\theta))$: since log is a monotonically increasing function, $\therefore$ $l(\theta)$ will increase if θ decreases.

for that lets analyse θ :

$$\theta \geq \|x\|$$

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

$\therefore$ the maximum likelihood estimate of θ is :

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

Ans1. The Approach mentioned in the question can only be used to solve a set of linear equations for eg in case of linear regression. In cases in which we need to solve a system of non linear equations gradient descent is needed. It is a more generic approach.

Also, the size of linear equations in linear regression maybe huge and we may have a memory constraint.

~~The problems in ML are also convex~~

We need to deal with convex problems as well, so gradients ensure that we get to the extrema.

Ans 2. In general a function approximation problem asks us to select a function from a defined class that closely matches a target function. We can use interpolation, extrapolation and curve fitting techniques of Numerical methods to approximate a function.

The major difference b/w ML and function approximation is that in ML the machine/model learns the pattern ~~of the~~ from the dataset provided and later predicts the output of an unknown input data.
We can use techniques such as Gaussian Naive Bayes, logistic Regression, Decision Trees etc to predict the outputs of an unknown data.

~~If we have all possible data that model ever seen them~~

In curve fitting we are often interested in parameters for a mathematical model based on a theory of cause &

effect underlying data, which may give random | systematic errors.

ML ~~go~~ gives a model by a task of discovering information.

If both of them are given all the data They can ever see then in function approx, we'll find a suitable best approximated function to the target function and no. of parameters will increase. The task will be huge and complexity will increase.

~~If~~ in case of machine learning, we can split the data ~~to~~ into ~~test~~ test and train and keep on improving our model by techniques of ML.

∴ They are different.