Theory Question

Ans.1 If we consider that the given number of parameters, n, is large, which is always the case in machine learning, then finding the minimum of the function involves solving a system of n equations. In the simplest case, this will in general involve inverting an nxn matrix. The standard matrix inversion algorithm is O(n^3) time. So when n is large, even linear regression can take a very long time to solve directly. And if the equations are not even linear, then the problem of finding an exact solution will be even more difficult, if not impossible. But gradient decent makes it an iterative problem and , makes sure that in the convex problems in machine learning, we get to the extrema.

Ans. 2 Machine learning is a method of data analysis that automates analytical model building. Here we are given a set of samples and we try to learn to build a target function using those samples as parameters. Also, the actual target function is also provided here to evaluate the accuracy. This training on seen data, forms the basis of the prediction target function to predict for the unseen data.
Where as, function approximation is a statistical technique to find a function fitting the given sample points. Here, no actual function is provided for the evaluation of the newly found function. All the future prediction for the unseen data, forms basis from this approximation only.

However, if we have all the data the model is expected to ever see, then both machine learning and function approximation behaves the same. Both techniques will be able to approximate the best function that can fit the given data.

PTO.

Ans.4

Page No.

Date

ans.4 Given, p.d.f. of the distribution $D=\{x_1, x_2 \dots x_n\}$

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

Likelihood function,

$$p(D|\theta) = \prod_{i=1}^{n} p(x;\theta)$$

$$= \prod_{i=1}^{n} \frac{1}{\pi^n \theta^{2n}} = \frac{1}{\pi^n} \times \prod_{i=1}^{n} \frac{1}{\theta^{2n}}$$

For maximum likelihood, we calculate the derivative of the log-likelihood,

$$\frac{d}{d\theta}\left(\ln p(D|\theta)\right) = \frac{d}{d\theta} \ln\left(\pi^{-n}\theta^{-2n}\right)$$

$$= \pi^{-n} \frac{d}{d\theta}\left(\ln \theta^{-2n}\right)$$

$$= \pi^{-n}\left(\frac{-2n}{\theta}\right) < 0$$

This implies that maximum likelihood is a decreasing function.

~~Since~~ whenever, $\|x\| > \theta$, $p(D|\theta) = 0$

$\therefore$ maximum likelihood is achieved at $\theta = x_n$

when given iids $\{x_1, x_2, \dots x_n\}$ are in increasing order.