B-TECH 4th SEMESTER MID TERM EXAMINATION

April 2021

Subject: Introduction to Data Science [CS 2205]

Date: 21/04/2021

Name: Abhiroop Mukherjee

Enrolment No.: 5105.19109

G-Suite ID: 510519109.abhirup@students.iiests.ac.in

No. of sheets uploaded: 8

Q1) a) → Data discretization is a process through which we convert continous data like height, weight, etc, to a finite set of intervals, and assign each interval with a data value

Eg ~~Weig~~ Height: 0m - 1m → short
1m - 1.5m → medium
1.5m - ∞ → tall

→ This basically reduces data ~~si~~ size, converting continous data into a set of values.

b) ~~Equal width discre~~
→ Converting to discrete data or bins simplifies function as we need to predict ~~ase~~ smaller things
→ This also avoids small fluctuations in data, which are mostly noise.

b) I) Equal width Discretization

→ All the bins have equal range width

Eg: Weight: 0 – ~~100~~ 50 kg : Underweight
50 – 100 kg : Average
100 – ~~200~~ 150 kg : Overweight

→ To put it mathematically:

Bins: [min + w], [min + 2w], ... [min + nw].

where $w = \frac{[max - min]}{n}$

n → no. of bins to make.

II) Equal-Frequency Discretization.

→ Bins are made such that every bins have ~~equal~~ approximately equal number of data points.

→ We have to select appropriate no. of bins to discretize data.

c) Main Challenge of Discretization

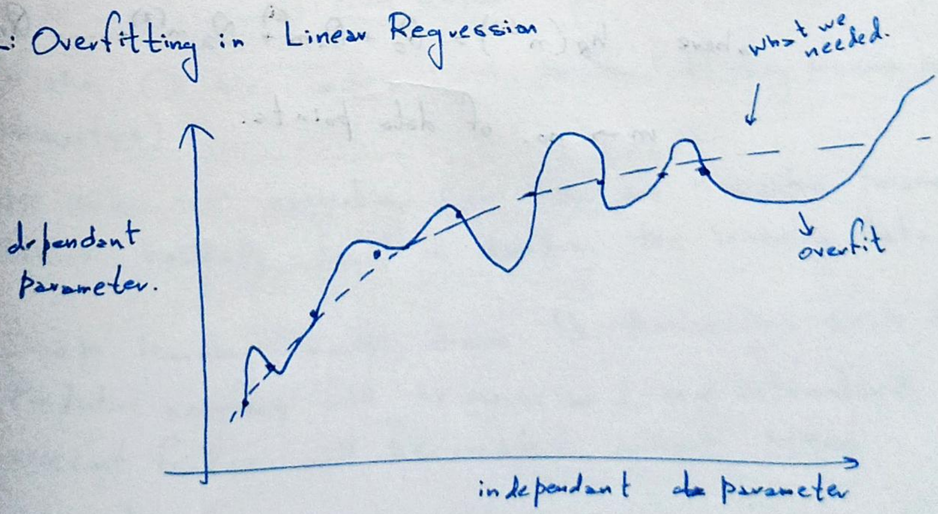
→ Discretization is basically a data reduction mechanism, hence by using Discretization, we are losing valuable data in trade for some avg/min/max of the bin. This trade is really risky.

→ ~~To to a~~ Also to avoid losing too much data, we need to find a good n, i.e no. of bins, to low n will give a lot of data loss, too high n will not give much data reduction.

~~Q3) a) Overfitting is a state of a model where it gets so focussed~~ ~~to reduce~~ ~~in training Accuracy that~~ ~~its~~ ~~it actually fails to generalize for the~~

Q3) a) Overfitting is a state of a model where it ~~learn~~ learns the sample ~~population~~ so much, that it fails to generalize parameters for the whole population.

Eg: Overfitting in Linear Regression



→ Here we see Training Loss is near zero, as the trained curve perfectly coincides with training result, but when compared to test data, the model performs horibly.

→ We need to avoid overfitting to best generalize the entire population using a sample test dataset.

b) Loss Function with L2 Regularization:

$$\cancel{Loss = Error(y,\hat{y}) + \lambda \sum_{i=1}^{N} w_i^2}$$

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} w_i^2$$

→ where λ is a very high constant

For Linear Regression:

$$\underset{(J(\theta))}{Loss} = \frac{1}{2n} \sum_{i=0}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 + \lambda \sum_{i=1}^{n} \theta_i^2$$

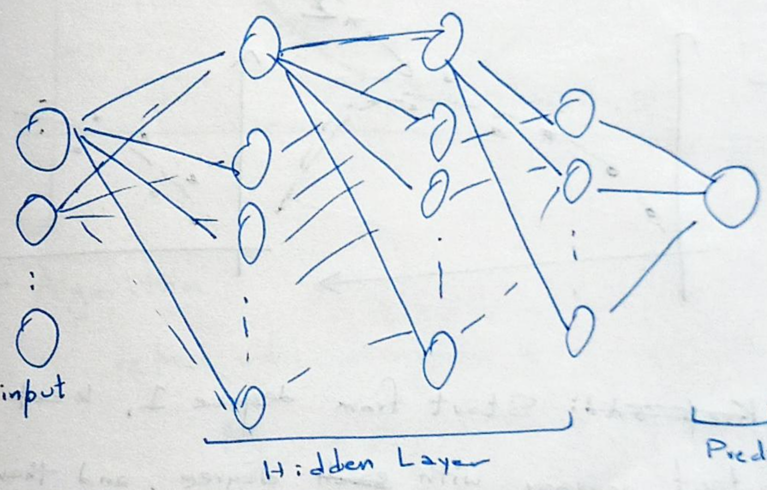
where $h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(1)} + \theta_2 x^{(2)} + \dots \theta_n x^{(n)}$

$m \to$ no. of data points.

c) → Deep Learning can ~~m~~ be made such to start with raw data, and using hidden layer in network, it will automatically create some feature from data, and using some more layers, predict the paramete required.

→ This ~~req~~ method required no manual feature extraction [although training speed will be incresed with features as input], as it can automatically create features as ~~require~~ required

General structure of a Multi Layer Deep Neural Network:



→ all the ○ are nodes (data processed by using previous layer parameters)

→ all the — connecting two nodes are trainable parameters which basically boost or dampen the incoming data.

→ Deep learning model trains it's Parameters such that Predictive accuracy will be maximized, and automatically relevant feature will be created without human intervention.

→ ~~This case of use ma sure~~

→ The training can be done through multiple epochs to increse training accuracy. [this can promote overfitting

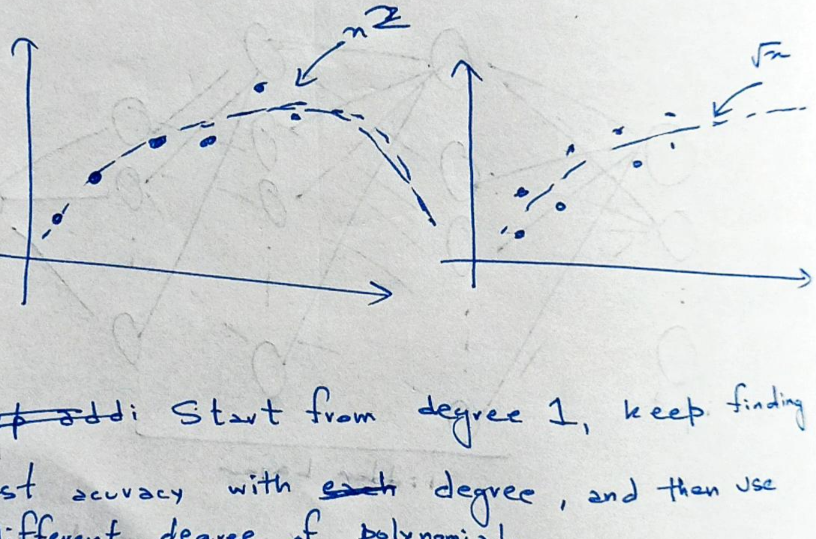
→ Due to ease of use and high accuracy, this model is good for predictive analysis.

Q6) a) We can decide degree of polynomial by using following ways:-

i) Plot a scatter plot of dataset, if possible

→ use scatter plot to make intelligent guesses about degree

Eg



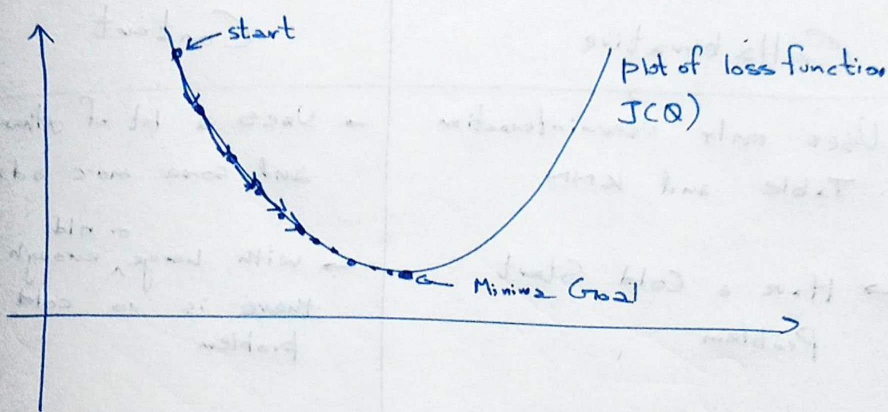
ii) ~~Keep addi~~ Start from degree 1, keep finding test accuracy with ~~each~~ degree, and then use different degree of polynomial,

→ Select the degree with maximum test accuracy as it best generalizes the whole population. [minimal overfitting]

b) Gradient Descent is an iterative process to minimize the loss of the model.

→ In this method, we iteratively update our ~~paramet~~ trainable parameter such that in each iteration, loss is minimized.

→ we use derivatives (or slopes) to iteratively go towards some minima [no idea if it's local or global minima]



→ Algorithm

repeat until convergence {

$$\theta_j := \theta_j - \alpha \left[\frac{\partial J(\theta)}{\partial \theta_j} \right]$$

}

→ Here we need to pick learning rate α

→ Too high α will lead to overshoot the iterative step, which ultimatly results in use never going to minima

→ Too low α will make this process go through a lot of iterations to reach minima.

→ Here convergence means that we don't see appreciable change in cost fn $J(\theta)$ through iteration, which means minima is reached.

c) → Collaborative Recommender System directly uses the ~~user into~~ user-item interaction table to predict recommendation, using KNN Algorithm

→ Content Based Approach also uses various other parameters like movie genre, cast, voice actors, songs, etc, and ~~use~~ feeds it into a model, which predicts whether a given user will likes it or not

Collaborative	Content
→ Uses only user-interaction Table and KNN	→ Uses a lot of other parameters and some more advanced model
→ Have a Cold Start Problem	→ with Large or old enough model, there is no cold start problem