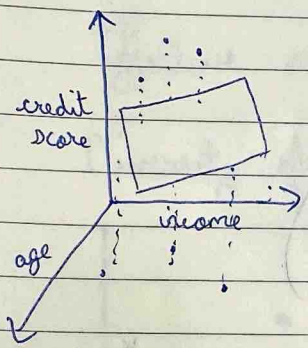
Lecture - 5 C8EcDd

Token No:- C8EcDd

Summary:-

At the beginning of lecture prof. gave a recap of the error measure, which will tell the risk factor associated to the error.

Then moving forward, we discussed about linear models for 3 learning problems which were classification Regression & Logistic regression. Then we discussed about if we have 2 attributes to determine the credit score of a user, we have to create a 3 dimensional representation.



This is how we will be creating plane which will act as threshold.

$$h(x) = w^T . x$$

If we get $w^T . x > 0$, the customer will be approved
If $w^T x < 0$ then the customer would be rejected.

Least squares linear regression:- In this we have to find a hypothesis with least error i.e the distance of data points from the hypothesis should be less. This could be calculated using the formula

$$E_{in}(h) = \frac{1}{N} \sum_{N=1} \left(h(x) - f(x)\right)^2$$

E_{in} → In sample error. $f(x)$ → data points.

$$E_{out}(h) = E_x\left[\left(h(x) - f(x)\right)^2\right]$$

Then we discussed about the dichotomy which represents correct classification of red and blue samples. We found that the effective number of hypothesis at most will be 2^N, this is for infinite Hypothesis set But we were only able to prove this for few values of sample set but for others we were getting exception.

Keeping all this in mind we were able to replace the value of Hypothesis set in Hoeffding bound eqn.

$$E_{out}(g) \leq E_{in}(g) + O\left(\sqrt{\frac{1}{N} \log \frac{m_H(N)}{\delta}}\right)$$

Our goal is to find a break point k such that

$$m_H(k) < 2^k.$$