Mid Semester Examination (6th Sem), 2022

Subject Name :→ Machine Learning.

Subject Code :→ IT 3205

Date of Examination :→ 12.03.2022

Name :→ Aniket Majhi.

Examination Roll Number :→ 510819019

G Suite ID :→ 510819019.aniket@students.iiests.ac.in

Number of sheets uploaded :→ 9.

with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with Experience E.

☐ Role of target function in designing a learning system,

In designing the learning system, the role of the target function is very importand because it actually determines how effective your learning system would be and also it works in solving / predicting values of the training examples.

(b.) A hypothesis h_j is more-general-than or equal to h_k if and only if any instance that satisfies h_j also satisfies h_k.

Let,

h_j and h_k be boolean-valued functions defined over X.

Then h_j is more general than or equal to h_k

if and only if

$$(\forall x \in X)\left[(h_k(x) = 1) \rightarrow (h_j(x) = 1)\right].$$

Example :→

h_1 = < value, ?, ?, ?, value >

h_2 = < value, ?, ?, ?, ? >

in the above example h_2 is more general than h_1.

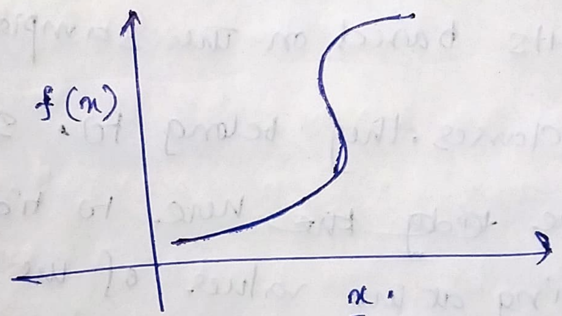
(4.)

(a.) Logistic regression :→ It is the appropriate regression analysis to conduct when the dependent variable is binary.

~~The equation~~

The objective function is given,

$$f_{w,b}(x) = \frac{1}{1+e^{-(wx+b)}}.$$



☐ optimization criterion :—

(b.)

Split in Decision Tree learning :→ A decision tree makes decisions by splitting nodes into sub nodes. This process is performed multiple times during the training until only the leaves node are left.

We know that in decision trees, the goal is to tidy the data. We split the data into two parts based on the samples together in the classes they belong to. So decision trees are ~~tidy the~~ here to tidy the dataset by looking at the values of the feature vector associated with each ~~other~~ data point and of the features the based on this value ~~the~~ of the features the decisions are made.

At each step, each branching, our target is to decrease the entropy, so this quantity is computed before each cut and also after the cut as well. If the entropy decrease then the cut is valid and we can process toward next step.

(6.)

(a.) Distance weighted nearest neighbour algorithm:-

We know that in K-nearest neighbour algorithm, we try to find out the K nearest neighbours of any new data point and based on the values we try to find out the value of the new data point. For classification we use the majority result of the K-nearest neighbors and for regression we use the average of the K nearest neighbours to find out the result.

But, ~~this process is not~~ what if we use and extra parameter with this result that is we give the weights based on their distances from the new data points. In this case the data points that are way far from the new data will get automatically get eliminated due to the weight factor.

In KNN,

$$\hat{f}(x_q) \leftarrow \underset{v \in V}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(v, f(x_i)).$$

in Distance weighted.

$$\hat{f}(x_q) \leftarrow \underset{v \in V}{\operatorname{argmax}} \sum_{i=1}^{k} w_i \delta(v, f(x_i))$$

weights

here, $w_i = \dfrac{1}{d(x_q, x_i)^2}$

$\downarrow$

more distance less weight factor.

(6.)

Curse of dimentionality :—

If we apply the KNN to a problem in which each instances is described by a large number of attributes but where only few of these attributes are relevant for the particular target function, in this case, instances that have identical values for the relevent attributes may be distant from one another.

It results the similarity metric used by the KNN will be misleading. The distance between neighbors will be dominated by the large number of irrelavent attributes. It is refered to as the 'Curse of dimentionality'.

- To remove the curse of dimentionality, there is one approach that is to weight each attribute differently while calculating the distance between two instances.

(2.)

(a) For classification task with missing inputs, the learning algorithm must learn a set of functions.
Those functions corresponds to classifying x with a different subset of its input missing.

The situation can be handled efficiently is by defining such a large set of functions is to learn a probability distribution over all of the relevant variables, then solve the classification task by marginalizing out the missing variables.

(b.)

Classification	Regression
(i) It is the task of predicting a discrete class level.	(i) It is the task of predicting a continuous quantity
(ii) It may predict a continuous value is in the form of a probability for a class label.	(ii) It may predict a discrete value, but the discrete value in the form of an integer quantity.

☐ If there is a several instances of the random vector given,

~~(7.4)~~ ~~Prioror pr~~

☐ Prioror probability :→

A prior probability is the probability that an observation will fall into a group before you collect the data. ~~the prior~~ is ~~a probability~~ It represents the uncertainty over θ before you have sampled. denoted by $\pi(\theta)$.

☐ Posterior probability :→

A posterior probability is the probability of assigning observation to groups given the data. The posterior probability distribution representing your uncertainty over θ after you have ~~ko~~ sampled data.

denoted by $\pi(\theta/x)$.

maximizes the entire posterior distribution

A map estimates is the mode of the posterior distribution.

☑ **ML** :→ The maximum likelihood estimate (MLE) of a parameter is the value of the parameter that maximizes the likelihood, where the likelihood is a function of the parameter and is actually equal to the probability of the data condition on the parameter.