Machine Learning

Seminar 2 Solution

[Q1 Sample Solution]

A classifier is a machine learning model that is used to discriminate different objects based on certain features.

[Q2 Sample Solution]

A categorical variable (also called qualitative variable) is a variable that can take on one of a limited, and usually fixed, number of possible values, assigning each individual or other unit of observation to a particular group or nominal category on the basis of some qualitative property.

We can use values 0, 1, 2, … to represent each category. For example, for the humidity variable, Normal is 0 and High is 1.

[Q3 Sample Solution]

The value range of y in the logistic function is (0, 1).

It ensures the 1/y and the natural log for the transformation of logistic function.

[Q4 Sample Solution]

The parameter values are found such that they maximise the likelihood that the process described by the model produced the data that were observed.

As explained, Logistic regression uses the Maximum Likelihood for parameter estimation. The logistic regression equation is given as

$$F(g(x)) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

The parameters to be estimated in the equation of a logistic regression are β vectors.

To estimate β vectors consider the N sample with labels either 0 or 1.

Mathematically, For samples labeled as '1', we try to estimate β such that the product of all probability p(x) is as close to 1 as possible. And for samples labeled as '0', we try to estimate β such that the product of all probability is as close to 0 as possible in other words (1 — p(x)) should be as close to 1 as possible.

The above intuition is represented as

$$for\ samples\ labelled\ as\ 1: \prod_{s\ in\ y_i=1} p(x_i)$$

$$for\ samples\ labelled\ as\ 0: \prod_{s\ in\ y_i=0} (1 - p(x_i))$$

x_i represents the feature vector for the i^{th} sample.

On combining the above conditions we want to find β parameters such that the product of both of these products is maximum over all elements of the dataset.

$$L(\beta) = \prod_{s\ in\ y_i=1} p(x_i) \ * \prod_{s\ in\ y_i=0} (1 - p(x_i))$$

*This function is the one we need to optimize and is called the **likelihood function.***

Now, We combine the products and take log-likelihood to simply it further

$$L(\beta) = \prod_{s} \left(p(x_i)^{y_i} \;\; * \; (1 - p(x_i))^{1-y_i} \right)$$

$$l(\beta) = \sum_{i=1}^{n} y_i \, log(p(x_i)) + (1 - y_i) \, log(1 - p(x_i))$$

$$where, \; l(\beta) \; is \; log - likelihood$$

Let's substitute p(x) with its exponent form

$$l(\beta) = \sum_{i=1}^{n} y_i \, log\left(\frac{1}{1 + e^{-\beta x_i}}\right) + (1 - y_i) \, log\left(\frac{e^{-\beta x_i}}{1 + e^{-\beta x_i}}\right)$$

$$l(\beta) = \sum_{i=1}^{n} y_i \left[log\left(\frac{1}{1 + e^{-\beta x_i}}\right) - log\left(\frac{e^{-\beta x_i}}{1 + e^{-\beta x_i}}\right)\right] + \; log\left(\frac{e^{-\beta x_i}}{1 + e^{-\beta x_i}}\right)$$

$$l(\beta) = \sum_{i=1}^{n} y_i \left[log\left(e^{\beta x_i}\right)\right] + log\left(\frac{e^{-\beta x_i}}{1 + e^{-\beta x_i}} \; * \; \frac{e^{\beta x_i}}{e^{\beta x_i}}\right)$$

$$l(\beta) = \sum_{i=1}^{n} y_i \beta x_i + log\left(\frac{1}{1 + e^{\beta x_i}}\right)$$

Now we end up with the final form of the log-likelihood function which is to be optimized and is given as

$$l(\beta) = \sum_{i=1}^{n} y_i \beta x_i - log\left(1 + e^{\beta x_i}\right)$$

[Q5 Sample Solution]

Root Node: Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.

Leaf Node: Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.

Branch/Sub Tree: A tree formed by splitting the tree.

Parent/Child node: The root node of the tree is called the parent node, and other nodes are called the child nodes.

[Q6 Sample Solution]