|  |  |
| --- | --- |
|  | ASSIGNMENT 6 |
|  |  |
|  | **Andrew-ID: parmenin**  DATA INFERENCE AND APPLIED MACHINE LEARNING (18-785)  11/21/22 |

**Niyomwungeri Parmenide ISHIMWE**

----------------------------------------------------------------------------------------------------------------

I, the undersigned, have read the entire contents of the syllabus for course 18-785 (Data

Inference and Applied Machine Learning) and agree with the terms and conditions of

participating in this course, including adherence to CMU's AIV policy.

Signature: **Niyomwungeri Parmenide ISHIMWE**

Andrew ID: **parmenin**

Full Name: **Niyomwungeri Parmenide ISHIMWE**

**The libraries used:**

1. import numpy as np
2. import pandas as pd

**QUESTION 1: NONLINEARITY**

* 1. It may be required to consider non-linear correlations between variables because there are some statistical situations where an independent variable and a dependent variable do not have a straight-line or direct relationship. In this situation, the rate of increase or decrease or the slope of the curve showing the relationship can change as one of the variables changes. In a nonlinear relationship, change in any of the inputs is not proportional to changes in the output. Because it is flexible, it can be useful to produce appropriate results when similar future situations or inputs are present, such as choices, categorization, or inferences, a more flexible nonlinear analysis is required. A nonlinear relationship forms a curve when it is plotted on a graph as opposed to a straight line when a linear relationship is presented. One popular example of a non-linear relationship is a relationship between weight loss and the amount of sport exercise a person does.
  2. Non-linear relationships can be modeled mathematically in various shapes like exponential, logarithmic, logistic, trigonometric, power functions, Lorenz curves, Gaussian, and so on. the following formula is the general equation of non-linear regression.

**Y = f (X, β) + ε** where **X** = a vector of p predictors, **β** = a vector of k parameters, **f (-)** = a known regression function, and **ε** = an error term [1]. It is applicable for forecasting, predicting, and financial modeling. It can be appropriate for agricultural research purposes because many crops and soil relationships require non-linear relationships to produce reliable results.

* 1. A parsimonious model is one that uses the fewest feasible predictor variables to provide the desired level of explanation or prediction. And so, a nonlinear model can be more parsimonious than a linear model because, in statistical modeling, there are occasions when a non-linear model with fewer parameters fits the data well but using a linear model would require many more parameters to get a decent fit [2].

Even while linear regression is simpler to use and understand, the types of curves it can fit are somewhat limited. There are situations when using fewer data won't allow it to fit the curve. Contrarily, nonlinear regression can fit many more varieties of curves, but it might be more difficult to determine the optimum fit and understand the significance of the independent variables. [3]. The mathematical formula for linear models as provided in the course notes can be written as: **y = a + bx + ε** where **a** is the intercept (also known as constant); **b** is the slope (indicates how y depends on x), and the **ε** is the model errors or residuals. From the previous question, the general equation for non-linear regression can be written as **Y = f (X, β) + ε**. This shows that a linear formula’s variable can only be raised to the power of 1 only which indicates that its slope is constant, while the one for non-linear can be raised to the power of 2 or more which indicates that its slope is variable [4].

* 1. According to the course notes, the characteristics that are typically preserved when generating surrogates are:
* Specifying a null hypothesis H0 describing a linear process and then generating several surrogate data sets according to H0 using Monte Carlo methods.
* Calculating a discriminating statistic for the original time series and all the surrogate sets.
* If the value of the statistic is significantly different for the original series than for the surrogate set, the null hypothesis is rejected, and non-linearity is assumed.

As listed in course notes, two of the surrogate techniques or methods are random shuffle and random phases. Random shuffle surrogates are generated by randomly shuffling the original data, obtaining the same distribution, and destroying linear correlations. Random phases surrogate data are generated by the inverse Fourier Transform of the amplitudes of the Fourier Transform of the original data with new (uniformly random) phases. This approach preserves the linear correlations in the data.)

* 1. Mathematically, information is defined as **I(p) = log(1/p) = - log(p)** and this definition satisfies three properties such as **I(p) ≥ 0**  to mean that information is a non-negative quantity, **I(1) =** 0to mean that events that always occur do not communicate information, and **I(p1 p2) = I(p1) + I(p2)** to mean thatinformation due to independent events is additive.

Entropy is a measure of disorder or expected surprise. it measures the average uncertainty in the value of the discrete-valued probability density. The entropy H(X) of a discrete random variable is defined in a mathematical way as **H = -Σk p(k) ln p(k)**, where p(k) is the probability of recurrences times, with k=1…, K.

Mutual information given as **I(x,y) = H(x) + H(y) − H(x,y)** is a metric that quantifies the amount of information that one variable provides about another variable and is typically measured in units called “bits”. A value of zero indicates no dependence and larger values imply the existence of a relationship (which might be linear or nonlinear).

The degree of regularity and unpredictability of variations in time series data can be measured using the approximation entropy, which is also used to estimate entropy. Entropy can be used to determine the regularity in heart rate.

Mutual information is a method for calculating the amount of information that one variable informs another. To choose the best features, one can compare the impact of each feature on the dependent variable using mutual information between each feature and the dependent variable. Mutual selection may be preferable to correlation since a feature may be less correlated with a dependent variable while still offering a wealth of information about the dependent variable.

**QUESTION 2: CLASSIFICATION USING TREES**

**2.1)** A decision tree is a type of flowchart that shows the reasoning behind a course of action. It is a type of algorithm used in data analytics that uses conditional "control" statements to categorize data. A decision tree has one node (or "node") at the beginning and branches (or "splits") in two or more directions from there. Until a conclusion is achieved, each branch contains a wide range of potential outcomes that can be attained by a variety of choices and arbitrary occurrences.

* **Nodes** are a representation of the potential characteristics associated with an event. The property representing the most information gain is the root node.
* The **branches** represent the values of the qualities, while the leaves represent the classes.

Sometimes, decision trees can get quite complex. In these circumstances, they can end up giving irrelevant information an excessive amount of weight. We can eliminate nodes using a process called "pruning" to get around this problem. Pruning is exactly what it sounds like: we remove any unneeded branches as they emerge from the tree.