|  |  |  |  |
| --- | --- | --- | --- |
| **MEM 838 Machine Learning**  **Midterm-Spring 2019-2020** | | | |
| **Name:**  **İSMET GÖKCAN** | **Surname:**  **İRİŞ** | | **Student No:**  **21501355** |
| **Signature:** | | **Attendance No:** | |

**Q1. (a): What is machine learning? (10 pts)**

**(b): What are the steps of supervised learning? (10 pts)**

Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop conventional algorithms to perform the needed tasks.

Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

In order to solve a given problem of supervised learning, one has to perform the following steps:

1.Determine the type of training examples. Before doing anything else, the user should decide what kind of data is to be used as a training set. In the case of handwriting analysis, for example, this might be a single handwritten character, an entire handwritten word, or an entire line of handwriting.

2.Gather a training set. The training set needs to be representative of the real-world use of the function. Thus, a set of input objects is gathered and corresponding outputs are also gathered, either from human experts or from measurements.

3.Determine the input feature representation of the learned function. The accuracy of the learned function depends strongly on how the input object is represented. Typically, the input object is transformed into a feature vector, which contains a number of features that are descriptive of the object. The number of features should not be too large, because of the curse of dimensionality; but should contain enough information to accurately predict the output.

4.Determine the structure of the learned function and corresponding learning algorithm. For example, the engineer may choose to use support vector machines or decision trees.

5.Complete the design. Run the learning algorithm on the gathered training set. Some supervised learning algorithms require the user to determine certain control parameters. These parameters may be adjusted by optimizing performance on a subset (called a validation set) of the training set, or via cross-validation.

6.Evaluate the accuracy of the learned function. After parameter adjustment and learning, the performance of the resulting function should be measured on a test set that is separate from the training set.Training data already trained.

**Q2. Let’s assume that we have a Toyota corolla prices along with a related car attribute.**

|  |  |  |
| --- | --- | --- |
| **No** | **KM** | **Price (1000$)** |
| **1** | **46000** | **13.5** |
| **2** | **72000** | **13.75** |
| **3** | **41700** | **13.95** |
| **4** | **48000** | **14.95** |
| **5** | **38500** | **13.75** |

**Use this training data to predict the price of a corolla car with 45000 KM. To do this, please use (a): linear regression and (b): quadratic regression. (50 pts)**

**LINEAR REGRESSION**

**(a\*x)+b=Y**

**. =**

**. = .X = C =**

**= \* =**

**\* =**

**Y =**

**( \* ) \* Y = =**

**F(x)=(a\*x)+b = F(45000) = (-0.0031 \* 45000) + 1.4129e+04 = 1.3990e+04**

**QUADRATIC REGRESSION**

**(a\*) + b\*x + c = Y**

**\* =**

**(\*X) = T =**

**=**

**\* =**

**( \* ) \* Y = =**

**(a\*) + b\*x + c = F(x) = F(45000) = 1.4169e+04**

**Q3. Let’s assume that we have (x,y)={(1,2),(3,2),(4,6)}, use piecewise regression to estimate the fitting line. Assume that the x-axis is divided between -1 to 5. (30 pts)**

**ⱷ(-2) + ( + f) ⱷ(-1) - ⱷ(0) = 0**

**2ⱷ(-1) - 2ⱷ(0) =0 ⱷ(-2) = ⱷ(0)**

**ⱷ(-1) + ( + f)** **ⱷ(0) -** **ⱷ(1) = 0**

**(h=1 x=0 y=0)**

**-ⱷ(-1) +2ⱷ(0) - ⱷ(1) = 0**

**ⱷ(0) + ( + f) ⱷ(1) -** **ⱷ(2) = 2µ**

**(h=1 x=1 y=2)**

**-ⱷ(0) + (2+****2µ) ⱷ(1) - ⱷ(2) = 2µ**

**ⱷ(1) + ( + f) ⱷ(2) -** **ⱷ(3) = 0**

**(h=1 x=2 y=0)**

**-ⱷ(1) + 2ⱷ(2) - ⱷ(3) = 0**

**ⱷ(2) + ( + f) ⱷ(3) - ⱷ(4) =** **2µ**

**(h=1 x=3 y=2)**

**- ⱷ(2) + (2+µ) ⱷ(3) - ⱷ(4) = 2µ**

**ⱷ(3) + ( + f) ⱷ(4) - ⱷ(5) = 6µ**

**(h=1 x=4 y=6)**

**-ⱷ(3) + (2+ⱷ)ⱷ(4) - ⱷ(5) = 6µ**

**ⱷ(4) + ( + f) ⱷ(5) - ⱷ(6) = 0**

**ⱷ(4)=ⱷ(6) -ⱷ(4) + ⱷ(5) = 0**

**× =**