國立臺北教育大學 110學年度第1學期

資料探勘 期末考試卷 日期： 2022/01/04

班級： 　　學號： 　　姓名：

1. 請根據附件1檔案，分別使用最近鄰居分類法(k-Nearest Neighbor Classification, KNN)、邏輯迴歸(Logistic Regression, LR)、神經網路(Artificial Neural Network, ANN)進行**分類處理**，前50筆資料為訓練集、後30筆資料為測試集，並以誤差平方和作為規劃求解目標，並將相關數值填入下表。

各模型相關條件如下：(45分)

* 1. KNN：1 ≤ 自變數權重(*wi*) ≤ 10、1 ≤ 半徑平方(*σ*2) ≤ 10，選項設定參數(最大時限: 300秒、反覆運算次數: 10次)。
  2. LR：-10 ≤ 自變數權重(*wi*) ≤ 10，選項設定參數(最大時限: 300秒、反覆運算次數: 10次)。
  3. ANN：-10 ≤ 自變數權重(*wi*, *θi*) ≤ 10、兩層神經層(第一層為4個神經元、第二層為3個神經元)、一層輸出層，選項設定參數(最大時限: 300秒、反覆運算次數: 10次)。

Ans：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | 訓練集 |  | 測試集 |  |
| 誤差平方和 | 誤判率 | 誤差平方和 | 誤判率 |
| KNN |  |  |  |  |
| LR |  |  |  |  |
| ANN |  |  |  |  |

1. 請根據附件2檔案，分別使用最近鄰居分類法(k-Nearest Neighbor Classification, KNN)、邏輯迴歸(Logistic Regression)、神經網路(Artifical Neural Network , ANN)進行**迴歸估算**，前70筆資料為訓練集、後30筆資料為測試集，並以誤差平方和作為規劃求解目標，並將相關數值填入下表。

各模型相關條件如下：(45分)

* 1. KNN：1 ≤ 自變數權重(*wi*) ≤ 10、1 ≤ 半徑平方(*σ*2) ≤ 10，選項設定參數(最大時限: 300秒、反覆運算次數: 10次)。
  2. LR：線性迴歸*y* = *w*0+*w*1\**x*1+*w*2\**x*2，-10 ≤ 自變數權重(*wi*) ≤ 10，選項設定參數(最大時限: 300秒、反覆運算次數: 10次)。
  3. ANN：-10 ≤ 自變數權重(*wi*, *θi*) ≤ 10、兩層神經層(第一層為2個神經元、第二層為2個神經元)、一層輸出層，選項設定參數(最大時限: 300秒、反覆運算次數: 10次)。

Ans：

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | 訓練集 | 測試集 |  |
| 誤差平方和 | 誤差平方和 | R2 |
| KNN |  |  |  |
| LR |  |  |  |
| ANN |  |  |  |

1. 請根據附件3檔案，使用*y* = *w*0+*w*1\**x*1+*w*2\**x*2+*w*3\**x*1*x*2+*w*4\**x*12+*w*5\**x*22迴歸公式計算變異分析表(ANOVA)，並將相關數值填入下表。(10分)

Ans：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 自由度 | 方差和 | 均方差 | F統計量 |
| 迴歸 |  |  |  |  |
| 殘差 |  |  |  |  |
| 總和 |  |  |  |  |