**B529 Assignment 2**

**Question 1.** A scoring function reported scores for 10 positive data points and 10 negative ones as follows:

|  |  |  |
| --- | --- | --- |
| No. | Label | Score |
| 1 | Positive | 25 |
| 2 | Positive | 21 |
| 3 | Positive | 20 |
| 4 | Positive | 19 |
| 5 | Positive | 18 |
| 6 | Positive | 17 |
| 7 | Positive | 14 |
| 8 | Positive | 13 |
| 9 | Positive | 10 |
| 10 | Positive | 5 |
| 11 | Negative | 18 |
| 12 | Negative | 15 |
| 13 | Negative | 13 |
| 14 | Negative | 12 |
| 15 | Negative | 10 |
| 16 | Negative | 8 |
| 17 | Negative | 5 |
| 18 | Negative | 4 |
| 19 | Negative | 3 |
| 20 | Negative | 3 |

Using python codes by computing FPR and TPR for each of the following score thresholds 3, 4, … , 25, and plot the ROC curve of the scoring function (25 points).

**Question 2.** The expression level of a gene follows a normal distribution . Four measurements of the gene reported the following expression values: 10, 13, 15, and 20. Use the maximum likelihood estimation method to estimate the value for parameter *θ*. Please both calculate manually and use python codes (25 points)

**Question 3.** A patient survive data (file patient1.csv in the folder) contains the information of 167 patients with two columns: the first column contains patient status (0 for live and 1 for dead) and the second column contains patient ages. Use the sklearn package in Python to do logistic regression using the column “age” as the input and column “status” as the output. You need to give the commands, report the final weights, and plot the logistic function (25 points).

**Question 4.** An immunotherapy data set (Immunotherapy.xlsx in the folder) contains information about wart treatment results using immunotherapy. The sheet “training data set” contains the information of 80 training data points and the sheet “test data set” contains the information of 10 test data points.

Use the keras package in Python to build a neural network with two hidden layers and an output layer, in which tanh is the activation function. You can decide the number of neurons in each layer. Use the training data set to train the network so that it can predict treatment results using the information of sex, age, time, number of warts, type, area, induration diameter. Give the commands to build and train the model and give prediction results for the test data (25 points).