**Question 1:**

1. Suppose a model is trained that helps in predicting whether the tumor is benign and malignant. After training the model, we apply it to a test set of 200 instances (also labelled) and the model produces the contingency table below. **[2\*2=4 marks]**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Predicted Class | |
|  |  | Malignant | Benign | |
| True Class | Malignant | 60 | 0 | |
| Benign | 120 | 20 | |

Compute the precision and recall of this model with respect to the “Malignant” class and with respect to the “Benign” class.

1. A machine learning algorithm has 3 hyperparameters with h1, h2, h3 number of different values. How many models do you need to build if you are doing 5 fold cross validation to choose the best model? **[2 marks]**
2. Suggest a scenario where you would prefer to use distance-weighted k-NN over k-NN? Explain your stance. **[2 marks]**
3. Do we need feature scaling in k-NN algorithm? Justify your answer using an example. **[2 marks]**

**Question 2:**

1. Consider a classification model with logistic regression and L2 regularization. Assuming that model is suffering from the problem of over-fitting, decreasing the value of regularization parameter helps in reduction of over-fitting. Is it True or False? Justify**. [2 marks]**
2. What type of regularization technique will you use for feature selection? Explain **[1 marks]**
3. Can we train logistic regression classifier for multi-class (assuming # of classes=4) classification? Explain. **[2 marks]**

**Question 3:**

Consider the following dataset:  **[5 marks]**

|  |  |  |  |
| --- | --- | --- | --- |
| **price** | **maintenance** | **Safety measures** | **Beneficial** |
| lowpriced | cheap | yes | yes |
| lowpriced | average | yes | yes |
| lowpriced | cheap | yes | no |
| lowpriced | excessive | no | no |
| fair | average | no | no |
| fair | average | no | yes |
| fair | excessive | yes | no |
| fair | excessive | yes | yes |
| overpriced | average | yes | yes |
| overpriced | excessive | yes | no |
| overpriced | excessive | yes | yes |

Classify the new instance given: “price = fair, maintenance = cheap, safety measures = yes”. Use Laplace smoothing only when needed to avoid zero probability.

**Question 4**:

Consider the following dataset. **[5 marks]**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sl# | Exercise | Blood pressure level | Follow good diet ? | Class |
| 1 | Yes | High | No | - ve |
| 2 | Yes | High | Yes | +ve |
| 3 | No | High | No | + |
| 4 | Moderate | High | No | + |
| 5 | Moderate | Normal | No | + |
| 6 | Moderate | Normal | Yes | - |
| 7 | No | Normal | Yes | + |
| 8 | Yes | High | No | - |
| 9 | Yes | Normal | No | + |
| 10 | Moderate | Normal | No | + |
| 11 | Yes | Normal | Yes | + |
| 12 | No | High | Yes | + |
| 13 | No | Normal | No | + |
| 14 | Moderate | High | Yes | - |

Use the above training dataset, Train decision tree classifier using information gain criteria. Find the attribute for the root node.

**Q 5.**

1. How does soft margin affect the classification SVM? **[2 marks]**
2. What is SVM optimization problem? What are the different inequality constraints in the optimization function? **[3 marks]**
3. What are kernels in SVM? Can the same optimization be used for non-linear data? Write and explain parameters of mathematical model of non-linear SVM. **[ 1+1+3=5 Marks]**

**Q 6. [5 Marks]**

1. Following are bagging rounds and decision stump rules for the input data. Using random forest algorithm and, find the accuracy of the algorithm. **[5 marks]**

f1 <= 1.5 -> Class = Yes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Input feature f1 | 1.1 | 1.3 | 1.4 | 1.6 | 1.8 | 1.9 |
| Class Label | Yes | Yes | No | No | Yes | Yes |

f1 <= 1.2 -> Class = Yes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Input feature f1 | 1.1 | 1.2 | 1.3 | 1.5 | 1.6 | 1.8 |
| Class Label | Yes | Yes | Yes | No | No | Yes |

f1 <= 1.6 -> Class = Yes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Input feature f1 | 1.1 | 1.3 | 1.5 | 1.6 | 1.8 | 1.9 |
| Class Label | Yes | Yes | No | No | Yes | Yes |