Question Number : 172 Question Id : 640653521165 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

Consider another linear classifier with $\mathbf{w}' = 3\mathbf{w}$.

How many points are misclassified by this new classifier?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

5

MLP

Section Id :	64065333939
Section Number :	11
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	25
Number of Questions to be attempted :	25
Section Marks :	50
Display Number Panel :	Yes
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	Yes
Maximum Instruction Time :	0

Sub-Section Number : 1

Sub-Section Id : 64065373969

Question Shuffling Allowed : No

Is Section Default? : null

Question Number : 173 Question Id : 640653521166 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 0

Question Label : Multiple Choice Question

THIS IS QUESTION PAPER FOR THE SUBJECT "DIPLOMA LEVEL : MACHINE LEARNING PRACTICE"

ARE YOU SURE YOU HAVE TO WRITE EXAM FOR THIS SUBJECT?
CROSS CHECK YOUR HALL TICKET TO CONFIRM THE SUBJECTS TO BE WRITTEN.

(IF IT IS NOT THE CORRECT SUBJECT, PLS CHECK THE SECTION AT THE TOP FOR THE SUBJECTS REGISTERED BY YOU)

Options :

6406531737074. ✔ YES

6406531737075. ✖ NO

Sub-Section Number : 2

Sub-Section Id : 64065373970

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 174 Question Id : 640653521167 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1 Selectable Option : 0

Question Label : Multiple Select Question

Ashok has to train a logistic regression model on a dataset with gradient descent approach. Which of the following solvers can he use?

Options :

6406531737076. ✖ newton-cg

6406531737077. ✖ lbfgs

6406531737078. ✖ liblinear

6406531737079. ✔ sag

6406531737080. ✔ saga

Sub-Section Number :	3
Sub-Section Id :	64065373971
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 175 Question Id : 640653521168 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider following code snippet:

```
from sklearn.utils.multiclass import type_of_target
import numpy as np
print(type_of_target(np.array([[0, 1], [1, 1]])))
print(type_of_target([1.0, 0.0, 3.0]))
print(type_of_target(['a', 'b', 'a']))
```

What will be the output of the above code snippet in the correct sequence?

Options :

6406531737081. ✔ 'multilabel-indicator'
'multiclass'
'binary'

6406531737082. ✖

'multiclass'
'multiclass'
'binary'

6406531737083. ✖ 'binary'
'multiclass'
'multilabel-indicator'

6406531737084. ✖ 'multilabel-indicator'
'continuous'
'binary'

Question Number : 176 Question Id : 640653521169 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Brajesh wrote following code snippet:

```
clf = Perceptron(max_iter=100,
                 random_state=1729)
```

He learnt that every time he calls fit() method on 'clf', the parameters learnt from the previous training session (i.e. previous call to 'fit()') are lost. What should he change in code so that this problem is removed?

Options :

6406531737085. ✔ Set 'warm_start=True'

6406531737086. ✖ Combine training data from different training sessions

6406531737087. ✖ Set 'retain_parameters=True'

6406531737088. ✖ This problem can not be solved.

Question Number : 177 Question Id : 640653521170 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider following variants of gradient descent algorithm:

1. full batch gradient descent
2. mini batch gradient descent
3. stochastic gradient descent

Which of the following variants of gradient descent can be implemented with SGDClassifier?

Options :

6406531737089. ✖ only 3

6406531737090. ✔ 1, 2 and 3

6406531737091. ✖ 1 and 2 only

6406531737092. ✖ 2 and 3 only

Question Number : 178 Question Id : 640653521174 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Suppose we have a multi-class classification problem with n classes. Which of the following methods require exactly n classifiers to solve this problem?

Options :

6406531737105. ✔ OneVsRestClassifier

6406531737106. ✖ OneVsOneClassifier

6406531737107. ✖ OutputCodeClassifier

6406531737108. ✖ MultiOutputClassifier

Question Number : 179 Question Id : 640653521176 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

What might be the possible output of the following code:

```
from sklearn.feature_extraction.text import CountVectorizer
corpus = ["Hello Hello World great"]
vectorizer = CountVectorizer()
X = vectorizer.fit_transform(corpus)
print(X.toarray())
```

Options :

6406531737113. ✔ [1 2 1]

6406531737114. ✖ [1 2 3]

6406531737115. ✖ [0 1 1]

6406531737116. ✖ [1 1 1]

Question Number : 180 Question Id : 640653521177 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider following code snippet:

```
import numpy as np
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import MinMaxScaler
from sklearn.linear_model import LinearRegression
steps = [
    ('imputer', SimpleImputer(missing_values=np.nan, strategy='mean')),
    ('scaler', MinMaxScaler()),
    ('model', LinearRegression())
]
pipe = Pipeline(steps = steps)
```

From the above code what pipe[1].fit_transform(X) does ? where X is a feature matrix

Options :

6406531737117. ✖ Replaces missing values with mean value of feature

6406531737118. ✔ Applies MinMaxScaling on the X

6406531737119. ✖ LinearRegression model fitting

6406531737120. ✖ None of these

Question Number : 181 Question Id : 640653521178 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

What might be the possible output of the following code:

```
from sklearn.metrics import mean_absolute_error
y_true = [3, -0.5, 2, 7]
y_pred = [2.5, 0.0, 2, 8]
mean_absolute_error(y_true, y_pred)
```

Options :

6406531737121. ✖ 0.00

6406531737122. ✔ 0.50

6406531737123. ✖ 0.72

6406531737124. ✖ 1.00

Question Number : 182 Question Id : 640653521179 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

What might be the possible output of the following code:

```
from sklearn.metrics import precision_score
y_true = [1,1,0,1,0,0,1,0,1]
y_pred = [0,1,0,1,0,1,1,1,1]
precision_score(y_true,y_pred)
```

Options :

6406531737125. ✖ 0.00

6406531737126. ✖ 0.33

6406531737127. ✔ 0.66

6406531737128. ✖ 0.99

Question Number : 183 Question Id : 640653521181 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Mention TRUE or FALSE: Feature scaling does not impact KNN model performance

Options :

6406531737133. ✖ TRUE

6406531737134. ✔ FALSE

Question Number : 184 Question Id : 640653521182 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

By using all features of a dataset accuracy score of 100% is achieved on the training set, but accuracy score of 70% on test set, which of the following statements is most relevant?

Options :

6406531737135. ✖ Model is underfitting

6406531737136. ✔ Model is overfitting

6406531737137. ✖ Nothing, the model is perfect

Question Number : 185 Question Id : 640653521185 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following code snippets correctly sets up a RandomizedSearchCV object to perform hyperparameter tuning on a RandomForestClassifier with the following parameters to be tested:

• Number of estimators: 50, 100, 150
• Maximum depth of the tree: 5, 10, 15
• Minimum number of samples to enable a split: 2, 4, 6

Options :

6406531737147. ✖

```
rfc = RandomForestClassifier(random_state=0)
param_distributions = {n_estimators: [50, 100, 150],
max_depth: [5, 10, 15],
min_samples_split: [2, 4, 6]}
randomized_search = RandomizedSearchCV(rfc,
param_distributions=param_distributions,
cv=5)
```

6406531737148. ✖

```
rfc = RandomForestClassifier(random_state=0)
param_distributions = ['n_estimators': [50, 100, 150],
'max_depth': [5, 10, 15],
'min_samples_split': [2, 4, 6]]
randomized_search = RandomizedSearchCV(rfc,
param_distributions=param_distributions,
cv=5)
```

```
rfc = RandomForestClassifier(random_state=0)
param_distributions = ['n_estimators': {50, 100, 150},
'max_depth': {5, 10, 15},
'min_samples_split': {2, 4, 6}]
randomized_search = RandomizedSearchCV(rfc,
param_distributions=param_distributions,
cv=5)
```

6406531737149. ✖

```
rfc = RandomForestClassifier(random_state=0)
param_distributions = {'n_estimators': [50, 100, 150],
'max_depth': [5, 10, 15],
'min_samples_split': [2, 4, 6]}
randomized_search = RandomizedSearchCV(rfc,
param_distributions=param_distributions, cv=5)
```

6406531737150. ✔

Question Number : 186 Question Id : 640653521186 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Decision Trees are prone to:

Options :

6406531737151. ✖ Low bias, low variance

6406531737152. ✖ High bias, low variance

6406531737153. ✔ Low bias, high variance

6406531737154. ✖ High bias, high variance

Question Number : 187 Question Id : 640653521189 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code. How many DecisionTreeClassifier models will be trained internally?

```
from sklearn.ensemble import BaggingRegressor
from sklearn.model_selection import GridSearchCV
param_grid = [{'max_depth':range(1, 20, 2)}]
gs = GridSearchCV(DecisionTreeClassifier(), param_grid, cv = 10)
gs.fit(X,y)
```

Options :

6406531737165. ✖ 1000

6406531737166. ✖ 20

6406531737167. ✖ 10000

6406531737168. ✔ 100

6406531737169. ✖ 90

Sub-Section Number : 4

Sub-Section Id : 64065373972

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 188 Question Id : 640653521171 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

The precision-recall curve

Options :

6406531737093. ✔ plots a graph with precision value on X-axis and recall value on Y-axis

6406531737094. ✔ computes precision-recall pairs for different probability thresholds

6406531737095. ✖ computes precision-recall pairs for one singular probability threshold

6406531737096. ✖ plots a graph with recall value on X-axis and precision value on Y-axis

Question Number : 189 Question Id : 640653521172 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction

Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Consider a classification dataset with 98% negative samples and 2% positive samples. A model is trained on this data, which of the following evaluation metrics are suitable for measuring effectiveness of this model:

Options :

6406531737097. ✖ accuracy

6406531737098. ✔ precision

6406531737099. ✔ recall

6406531737100. ✔ F-1 score

Question Number : 190 Question Id : 640653521173 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Consider following code snippet:

```
from sklearn.naive_bayes import MultinomialNB
estimator  = MultinomialNB()
estimator.fit(X, y)
```

where X and y are training data.

Options :

6406531737101. ✖ MultinomialNB is best suited when feature matrix X contains text data and not the word counts.

6406531737102. ✔ MultinomialNB is best suited when feature matrix X contains word counts for text data.

6406531737103. ✖ The MultinomialNB classifier is suitable for classification with continuous features.

6406531737104. ✖ None of these

Question Number : 191 Question Id : 640653521175 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following option(s) are correct regarding regularization?

Options :

6406531737109. ✔ It is a technique used to minimize the adjusted loss function and avoid overfitting.

6406531737110. ✖ It increases the bias and variance of the training model

6406531737111. ✔ Elastic net regularization is a combination of L1 and L2 regularization both.

6406531737112. ✖ It controls the number of passes a training dataset takes in an algorithm.

Question Number : 192 Question Id : 640653521180 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following statements are true? (Multiple options may be correct.)

Options :

6406531737129. ✔ KNN models with low values of K produces complex decision boundaries.

6406531737130. ✔ KNN models with high values of K produces smooth decision boundaries.

6406531737131. ✖ In KNN models K does not impact the decision boundaries.

6406531737132. ✖ None of these

Question Number : 193 Question Id : 640653521184 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Fill in the missing parameter value in the following estimator that can be used to classify the data

```
from sklearn.svm import SVC
clf = SVC(kernel = ______)
clf.fit(X, y)
```

Options :

6406531737143. ✔ 'poly',

6406531737144. ✖ 'lasso'

6406531737145. ✔ 'rbf',

6406531737146. ✖ 'scale'

Question Number : 194 Question Id : 640653521187 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following is/are correct?

Options :

6406531737155. ✖ Decision trees are prone to underfitting.

6406531737156. ✖ By increasing the 'max_depth' parameter in 'DecisionTreeClassifier', the tree is likely to underfit

6406531737157. ✖ By increasing the 'min_samples_leaf' parameter in 'DecisionTreeClassifier', the tree is likely to overfit.

6406531737158. ✖ By increasing the 'min_samples_split' parameter in 'DecisionTreeClassifier', the tree is likely to overfit.

6406531737159. ✖ By increasing the 'ccp_alpha' parameter in 'DecisionTreeClassifier', the tree is likely to overfit.

6406531737160. ✔ None of these

Sub-Section Number : 5

Sub-Section Id : 64065373973

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 195 Question Id : 640653521183 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Which of the following options are true for regularization parameter C in sklearn.svm.SVC ?

Options :

6406531737138. ✖ Large value of the regularization parameter C will overfit the training set and complex decision boundaries will form.

6406531737139. ✔ Large value of the regularization parameter C will underfit the training set and smooth decision boundaries will form.

6406531737140. ✔ Small value of the regularization parameter C will overfit the training set and complex decision boundaries will form.

6406531737141. ✖ Small value of the regularization parameter C will underfit the training set and smooth decision boundaries will form.

6406531737142. ✖ None of these

Question Number : 196 Question Id : 640653521190 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Consider the following block of code:

```
from sklearn.datasets import load_breast_cancer
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
X,y = load_breast_cancer(as_frame = True,
                         return_X_y = True)
X_train,X_test,y_train,y_test = train_test_split(X,y,
                                                 test_size = 0.2,
                                                 random_state = 1)
clf = DecisionTreeClassifier(min_samples_split = 5,
                             min_samples_leaf = 3,
                             random_state = 5)
clf.fit(X_train, y_train)
print(clf.score(X_test, y_test))
```

In which of the following scenarios, the split will NOT be made at node N?

Options :

6406531737170. ✖ Number of samples at node N = 10. If it is split, it will result in 4 nodes in the left child and 6 nodes in the right child.

6406531737171. ✖ Number of samples at node N = 6. If it is split, it will result in 3 nodes in the left child and 3 nodes in the right child.

6406531737172. ✔ Number of samples at node N = 12. If it is split, it will result in 2 nodes in the left child and 10 nodes in the right child.

6406531737173. ✔ Number of samples at node N = 4. If it is split, it will result in 3 nodes in the left child and 1 node in the right child.

Sub-Section Number :	6
Sub-Section Id :	64065373974
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 197 Question Id : 640653521188 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Consider the following code with respect to some feature matrix X and target vector y:

```
from sklearn.datasets import load_wine
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split

X_train,X_test,y_train,y_test = train_test_split(X,y,
                                                 test_size = 0.10,
                                                 random_state = 12)

clf = DecisionTreeClassifier(max_depth = 6,
                             min_samples_split = 2,
                             min_samples_leaf=3,
                             random_state = 81)

clf.fit(X_train, y_train)
print(clf.score(X_train, y_train))
```

Assume that the output of the above code is 0.852. If we increase the value of the parameter 'max_depth', which of the following is more likely to happen?:

Options :

6406531737161. ✔ The output score is likely to increase.

6406531737162. ✖ The output score is likely to decrease.

6406531737163. ✖ The change in 'max_depth' is not likely to have any effect on the output.

6406531737164. ✖ If we increase the value of 'max_depth' beyond 6, the code will throw an error, as the max_depth can not be more than the product of 'min_samples_split' and 'min_samples_leaf'.

BDM

Section Id :	64065333940
Section Number :	12
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	11
Number of Questions to be attempted :	11
Section Marks :	15
Display Number Panel :	Yes
Group All Questions :	No