Microsoft

DP-100

Designing and Implementing a Data Science Solution on Azure?

http://killexams.com/pass4sure/exam-detail/DP-100
**QUESTION:** 103

DRAG DROP

You have a model with a large difference between the training and validation error values. You must create a new model and perform cross-validation.

You need to identify a parameter set for the new model using Azure Machine Learning Studio. Which module you should use for each step? To answer, drag the appropriate modules to the correct steps. Each module may be used once or more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

**NOTE:** Each correct selection is worth one point.

<table>
<thead>
<tr>
<th>Modules</th>
<th>Step</th>
<th>Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-Class Boosted Decision Tree</td>
<td>Define the parameter scope</td>
<td>Split Data</td>
</tr>
<tr>
<td>Partition and Sample</td>
<td>Define the cross-validation settings</td>
<td>Partition and Sample</td>
</tr>
<tr>
<td>Tune Model Hyperparameters</td>
<td>Define the metric</td>
<td>Two-Class Boosted Decision Tree</td>
</tr>
<tr>
<td>Split Data</td>
<td>Train, evaluate, and compare</td>
<td>Tune Model Hyperparameters</td>
</tr>
</tbody>
</table>

**Answer:**
Exhibit

**Explanation:**
Box 1: Split data
Box 2: Partition and Sample
Box 3: Two-Class Boosted Decision Tree Box 4: Tune Model Hyper parameters
Integrated train and tune: You configure a set of parameters to use, and then let the module iterate over multiple combinations, measuring accuracy until it finds a "best" model. With most learner modules, you can choose which parameters should be changed during the training process, and which should remain fixed.
We recommend that you use Cross-Validate Model to establish the goodness of the model given the specified parameters. Use Tune Model Hyperparameters to identify the optimal parameters.

References:

QUESTION: 104
HOTSPOT
You are using C-Support Vector classification to do a multi-class classification with an unbalanced training dataset. The C-Support Vector classification using Python code shown below:

```python
from sklearn.svm import SVC
import numpy as np
svc = SVC(kernel='linear', class_weight='balanced', C=1.0, random_state=0)
model1 = svc.fit(X_train, y)
```

You need to evaluate the C-Support Vector classification code.
Which evaluation statement should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

**Answer:**
Exhibit
**Explanation:**
Box 1: Automatically adjust weights inversely proportional to class frequencies in the input data. The “balanced” mode uses the values of y to automatically adjust weights inversely proportional to class frequencies in the input data as \( \frac{n_{\text{samples}}}{n_{\text{classes}} \times \text{np.bincount}(y)} \).

Box 2: Penalty parameter
Parameter: \( C : \) float, optional (default=1.0) Penalty parameter \( C \) of the error term.

**References:**

**QUESTION:**
You are building a machine learning model for translating English language textual content into French language textual content. You need to build and train the machine learning model to learn the sequence of the textual content. Which type of neural network should you use?

A. Multilayer Perceptions (MLPs)
B. Convolutional Neural Networks (CNNs)
C. Recurrent Neural Networks (RNNs)
D. Generative Adversarial Networks (GANs)

**Answer:** C

**Explanation:**
To translate a corpus of English text to French, we need to build a recurrent neural network (RNN). Note: RNNs are designed to take sequences of text as inputs or return sequences of text as outputs, or both. They’re called recurrent because the network’s hidden layers have a loop in which the output and cell state from each time step become inputs at the next time step. This
recurrence serves as a form of memory. It allows contextual information to flow through the network so that relevant outputs from previous time steps can be applied to network operations at the current time step.

References:
https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571

**QUESTION: 106**
You create a binary classification model. You need to evaluate the model performance. Which two metrics can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. relative absolute error  
B. precision  
C. accuracy  
D. mean absolute error  
E. coefficient of determination

**Answer:** B, C

**Explanation:**
The evaluation metrics available for binary classification models are: Accuracy, Precision, Recall, F1 Score, and AUC.  
Note: A very natural question is: ‘Out of the individuals whom the model, how many were classified correctly (TP)?’
This question can be answered by looking at the Precision of the model, which is the proportion of positives that are classified correctly.

References:

**QUESTION: 107**
You use the Two-Class Neural Network module in Azure Machine Learning Studio to build a binary classification model. You use the Tune Model Hyperparameters module to tune accuracy for the model. You need to select the hyperparameters that should be tuned using the Tune Model Hyperparameters module. Which two hyperparameters should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.
A. Number of hidden nodes
B. Learning Rate
C. The type of the normalizer
D. Number of learning iterations
E. Hidden layer specification

**Answer:** D, E

**Explanation:**
D: For Number of learning iterations, specify the maximum number of times the algorithm should process the training cases.
E: For Hidden layer specification, select the type of network architecture to create. Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers.

**References:**

**QUESTION:** 108

HOTSPOT
You are evaluating a Python NumPy array that contains six data points defined as follows: data = [10, 20, 30, 40, 50, 60]
You must generate the following output by using the k-fold algorithm implantation in the Python Scikitlearn machine learning library:
train: [10 40 50 60], test: [20 30]
train: [20 30 40 60], test: [10 50]
train: [10 20 30 50], test: [40 60]
You need to implement a cross-validation to generate the output.
How should you complete the code segment? To answer, select the appropriate code segment in the dialog box in the answer area.
NOTE: Each correct selection is worth one point.
```python
from numpy import array
from sklearn.model_selection import KMeans, kfold, CrossValidation, ModelSelection

data = array([[10, 20, 30, 40, 50, 60]])
kfold = Kfold(n_splits=3, shuffle=True, random_state=1)

for train, test in kFold, split(data):
    print('train: %s, test: %s' % (data[train], data[test]))
```

**Answer:**
Exhibit

**Explanation:**
Box 1: k-fold
Box 2: 3
K-Folds cross-validator provides train/test indices to split data in train/test sets. Split dataset into k consecutive folds (without shuffling by default).
The parameter n_splits (int, default=3) is the number of folds. Must be at least 2. Box 3:
Example:

```python
>>> from sklearn.model_selection import KFold
>>> X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]])
>>> y = np.array([1, 2, 3, 4])
>>> kf = KFold(n_splits=2)
>>> kf.get_n_splits(X) 2
>>> print(kf)
KFold(n_splits=2, random_state=None, shuffle=False)
>>> for train_index, test_index in kf.split(X):
...    print("TRAIN:", train_index, "TEST:", test_index)
...    X_train, X_test = X[train_index], X[test_index]
...    y_train, y_test = y[train_index], y[test_index]
TRAIN: [2 3] TEST: [0 1]
TRAIN: [0 1] TEST: [2 3]
```

References:

**QUESTION:** 109

You create a binary classification model by using Azure Machine Learning Studio. You must tune hyperparameters by performing a parameter sweep of the model. The parameter sweep must meet the following requirements:
- iterate all possible combinations of hyperparameters
- minimize computing resources required to perform the sweep
- You need to perform a parameter sweep of the model. Which parameter sweep mode should you use?

A. Random sweep  
B. Sweep clustering  
C. Entire grid  
D. Random grid  
E. Random seed

**Answer:** D

**Explanation:**
Maximum number of runs on random grid: This option also controls the number of iterations over a random sampling of parameter values, but the values are not generated randomly from the specified range; instead, a matrix is created of all possible combinations of parameter values and a random sampling is taken over the matrix. This method is more efficient and less prone to regional oversampling or undersampling.
If you are training a model that supports an integrated parameter sweep, you can also set a range of seed values to use and iterate over the random seeds as well. This is optional, but can be useful for avoiding bias introduced by seed selection.

Incorrect Answers:
B: If you are building a clustering model, use Sweep Clustering to automatically determine the optimum number of clusters and other parameters.
C: Entire grid: When you select this option, the module loops over a grid predefined by the system, to try different combinations and identify the best learner. This option is useful for cases where you don't know what the best parameter settings might be and want to try all possible combination of values.
E: If you choose a random sweep, you can specify how many times the model should be trained, using a random combination of parameter values.

References:

**QUESTION:** 110
**HOTSPOT**
You are analyzing the asymmetry in a statistical distribution.
The following image contains two density curves that show the probability distribution of two datasets.

Use the drop-down menus to select the answer choice that answers each question based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.
Answer:
Exhibit

Explanation:
Box 1: Positive skew
Positive skew values means the distribution is skewed to the right.
Box 2: Negative skew
Negative skewness values mean the distribution is skewed to the left.

References:
For More exams visit https://killexams.com/vendors-exam-list

Kill your exam at First Attempt....Guaranteed!