Score: **47%**

No. of questions: **8**

Correct answer: **3**

Incorrect answer: **5**

Show incorrect attempt only ⬤

Question 1 ⊗   1 Mark

Why do we need a validation set?

| A | To make sure the model does not overfit on the training data. | ⊗ |

| B | To make sure the model does not overfit on the test data. |

| C | To make sure the model does not overfit on the training and test data. |

Correct Answer: B. To make sure the model does not overfit on the test data.

Validation data helps in achieving the best model with better generalization capability by overcoming the problem of overfitting.

Question 2 ⊗   2 Marks

For a given dataset, we have 2014 rows and 19 features. We can divide the dataset like this:-

◄ Prev
Training - 30% Quiz
Validation - 30% (machine-learning/screen/6184)

Next ►
Video
(machine-learning/screen/6185)

Can we consider this distribution to be good?

| A | Yes | ⊗ |

| B | No |

Correct Answer: B. No

Generally, the maximum proportion of dataset is preserved for training the model whereas the minimum proportion of dataset is preserved for validation and test. In this case, we have maximum proportion for the test data which is not a good distribution.

Question 3 ✅

3 marks

What is the right sequence for performing a K fold cross validation?

1. Pick a group as a holdout
2. Predict and evaluate on the holdout
3. Shuffle the dataset randomly
4. Take the remaining groups as training and fit a model
5. Split the dataset into k groups
6. Repeat the procedure with every group

| A | 1, 2, 3, 4, 5, 6 |

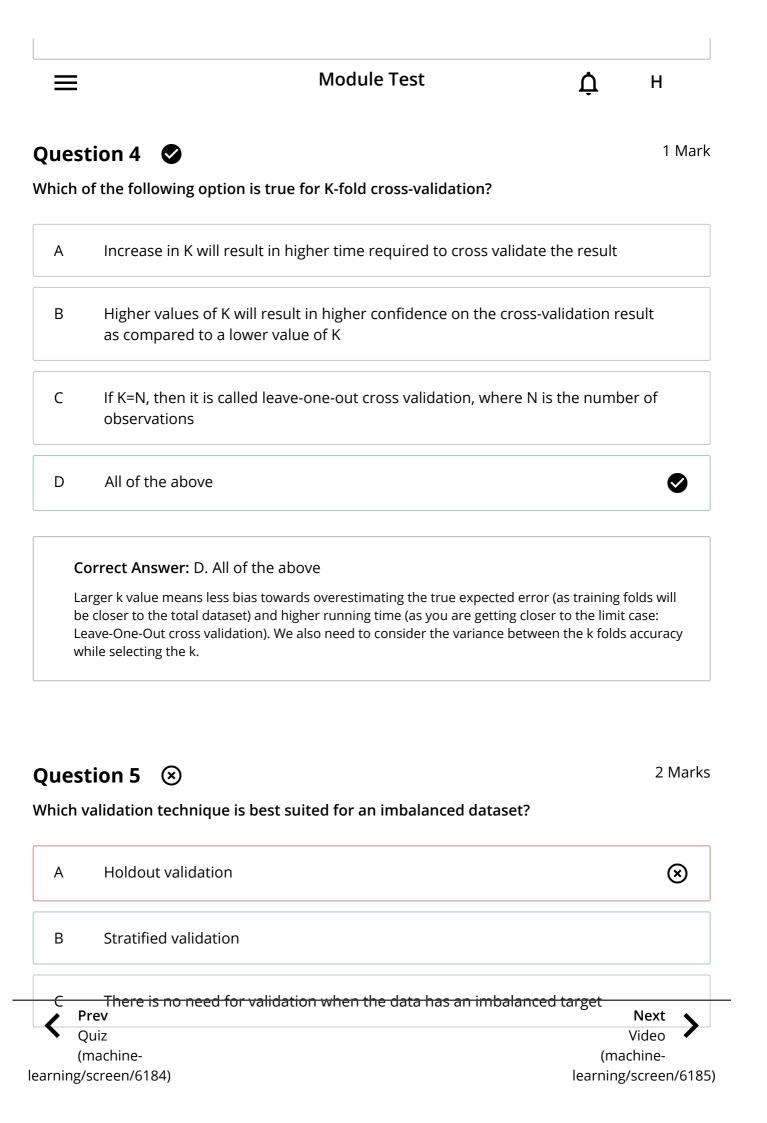| B | 6, 5, 4, 3, 2, 1 |

| C | 3, 5, 1, 4, 2, 6 | ✅ |

| D | 6, 2, 4, 1, 5, 3 |

Correct Answer: C. 3, 5, 1, 4, 2, 6

In k cross fold validation, initially we shuffle dataset randomly and split the dataset into k folds. The model is trained on k-1 folds and hold out fold is used for validation. This is repeated for k folds. Hence, the right sequence would be 3->5->1->4->2->6.

Question 4 ✅   1 Mark

Which of the following option is true for K-fold cross-validation?

A	Increase in K will result in higher time required to cross validate the result

B	Higher values of K will result in higher confidence on the cross-validation result as compared to a lower value of K

C	If K=N, then it is called leave-one-out cross validation, where N is the number of observations

D	All of the above ✅

Correct Answer: D. All of the above

Larger k value means less bias towards overestimating the true expected error (as training folds will be closer to the total dataset) and higher running time (as you are getting closer to the limit case: Leave-One-Out cross validation). We also need to consider the variance between the k folds accuracy while selecting the k.

Question 5 ⊗   2 Marks

Which validation technique is best suited for an imbalanced dataset?

A	Holdout validation ⊗

B	Stratified validation

C	There is no need for validation when the data has an imbalanced target

‹ Prev
Quiz
(machine-learning/screen/6184)

Next ›
Video
(machine-learning/screen/6185)

Correct Answer: B. Stratified validation

Stratification is a process of partitioning the data such that class distribution of the training data is preserved. It is helpful for class imbalance problems since the validation data must mimic the training data. Hence, option B is correct.

Question 6 ✅

3 marks

For a given dataset with n data points, K-fold cross-validation and leave one out validation are the same, when:

A	k=n	✅

B	k=N-1

C	k=n/2

D	k=2*n

Correct Answer: A. k=n

In leave one out validation, the model is trained on all data except one point and prediction is made for that point. In K cross fold validation, the model is trained on k-1 folds and hold out fold is used for evaluation. This process is repeated for k folds. When k=n, the K-fold cross-validation and leave one out validation are the same.

Question 7 ⊗

1 Mark

What is the underlying issue with the simple/stratified holdout validation technique?

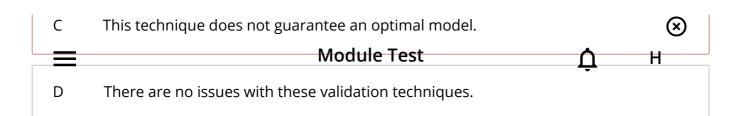| A	It may generate an uneven distribution.

B	We need to put aside some data.

‹ **Prev**
Quiz
(machine-learning/screen/6184)

Next ›
Video
(machine-learning/screen/6185)

C This technique does not guarantee an optimal model. ⊗

D There are no issues with these validation techniques.
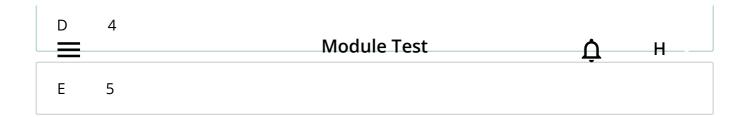
Correct Answer: B. We need to put aside some data.

During hold out validation technique, the dataset is split into train set as well as the validation set. The model is trained only on the train set and the validation set is never used to train the model. Hence, option b is correct.

Question 8 ⊗   2 Marks

Given below is a scenario for training error (TE) and Validation error (VE) for a machine learning algorithm M1. You want to choose a hyperparameter (H) based on TE and VE:

H	TE	VE
1	105	90
2	200	85
3	250	96
4	105	85
5	300	100

Which value of H will you choose based on the above table?

A   1   ⊗

B   2

C   3

‹ **Prev**
Quiz
(machine-
learning/screen/6184)

Next ›
Video
(machine-
learning/screen/6185)

Correct Answer: D. 4

Based on the data provided in the table, option D seems the best as it has the best sync between training error and validation error.

Prev
Quiz
(machine-
learning/screen/6184)

Next
Video
(machine-
learning/screen/6185)