

QUIZ:- 09

Started on	Montag, 10 Januar 2022, 10:37
State	Finished
Completed on	Dienstag, 11 Januar 2022, 10:46
Time taken	1 day
Marks	7.33/8.00
Grade	3.67 out of 4.00 (92%)

Question 1

Complete

Mark 2.00 out of 2.00

🚩 Flag question

Which of the following statements about Machine Learning (ML) Pipelines is true?

Select one or more:

- ☐ Good generalization performance of a pipeline requires only tuning of the feature extraction hyperparameters.
- ☒ Using ML Pipelines APIs can require additional implementation overhead to integrate custom feature extraction code into a pipeline.
- ☒ ML Pipelines consist of feature extractors and ML models that take the extracted features and make a prediction.

Question 2

Complete

Mark 2.00 out of 2.00

🚩 Flag question

Which of the following statements is true for n-gram bag-of-word features.

Select one or more:

- ☒ n-gram features are the frequencies of token sequences of length n in a text.
- ☐ If an n-gram feature vector accounts for all words in a language, then most word counts in such a vector computed on a text in that language will be between 1 and 10.
- ☒ Denoting the size of the vocabulary as V the memory consumption of n-gram features can become (constant factor for storing a token) times V^n

Question 3

Complete

Mark 2.00 out of 2.00

🚩 Flag question

Consider the following data set

`[['large'], ['small'], ['medium'], ['small'], ['large']]`

Select one or more:

- ☐ The one-hot encoded representation of the data is

```
array([[1., 0.],
       [0., 0.],
       [0., 1.],
       [0., 0.],
       [1., 0.]])
```

- ☒ The one-hot encoded representation of the data is

```
array([[1., 0., 0.],
       [0., 0., 1.],
       [0., 1., 0.],
       [0., 0., 1.],
       [1., 0., 0.]])
```

- ☐ The one-hot encoded representation of the data is

```
array([[1., 0., 0.],
       [0., 0., 1.]])
```

Question 4

Complete

Mark 1.33 out of 2.00

Flag question

Consider the data set

```
x = ['black', 'yellow', 'black', 'orange', 'green']
```

Which of the following programs computes a one-hot encoding of the data?

Note that the convention of columns referring to features and rows to data points can be ignored here.

Select one or more:

☐ import numpy as np

```
unique_items = list(set(x))
one_hot = [[1 if ui == w else 0 for ui in x] for w in unique_items]

np.array(one_hot)
```

☒ import numpy as np

```
items = []
for i in x:
    if i not in items:
        items += [i]

one_hot = []
for i in x:
    this_one_hot = []
    for ui in items:
        if i == ui:
            this_one_hot.append(1)
        else:
            this_one_hot.append(0)
    one_hot.append(this_one_hot)

np.array(one_hot)
```

☒ import numpy as np

```
unique_items = list(set(x))
one_hot = [[1 if ui == w else 0 for ui in unique_items] for w in x]

np.array(one_hot)
```

☐ import numpy as np

```
items = []
for i in x:
    if i not in items:
        items += [i]

one_hot = []
for i in x:
    this_one_hot = []
    for ui in items:
        if i == ui:
            this_one_hot.append(0)
        else:
            this_one_hot.append(1)
    one_hot.append(this_one_hot)

np.array(one_hot)
```