Universitetet i Stavanger

CANDIDATE

3256

TEST

DAT540 1 Introduksjon til datavitenskap

Subject code	DAT540
Evaluation type	Skriftlig eksamen
Test opening time	06.12.2022 09:00
End time	06.12.2022 12:00
Grade deadline	27.12.2022 23:59
PDF created	06.12.2022 12:18

Section 1

Question	Question title	Question type
ℹ	Info - dat540.22H	Information or resources
1	dat540.22h.04	Multiple Choice
2	dat540.22h.04	Multiple Choice
3	dat540.22h.04	Multiple Choice
4	dat540.22h.04	Multiple Choice
5	dat540.22h.04	Multiple Choice
6	dat540.22h.04	Multiple Choice
7	dat540.22h.04	Multiple Choice
8	dat540.22h.04	Multiple Choice
9	dat540.22h.04	Multiple Choice
10	dat540.22h.03	Multiple Choice
11	dat540.22h.03	Multiple Choice
12	dat540.22h.03	Multiple Choice
13	dat540.22h.03	Multiple Choice
14	dat540.22h.03	Multiple Choice
15	dat540.22h.03	Multiple Choice
16	dat540.22h.03	Multiple Choice
17	dat540.22h.03	Multiple Choice
18	dat540.22h.03	Multiple Choice

19	dat540.22h.03	Multiple Choice
20	dat540.22h.02	Multiple Choice
21	dat540.22h.02	Multiple Choice
22	dat540.22h.02	Multiple Choice
23	dat540.22h.02	Multiple Choice
24	dat540.22h.02	Multiple Choice
25	dat540.22h.02	Multiple Choice
26	dat540.22h.02	Multiple Choice
27	dat540.22h.02	Multiple Choice
28	dat540.22h.02	Multiple Choice
29	dat540.22h.02	Multiple Choice
30	dat540.22h.01	Multiple Choice
31	dat540.22h.01	Multiple Choice
32	dat540.22h.01	Multiple Choice
33	dat540.22h.01	Multiple Choice
34	dat540.22h.01	Multiple Choice
35	dat540.22h.01	Multiple Choice
36	dat540.22h.01	Multiple Choice
37	dat540.22h.01	Multiple Choice
38	dat540.22h.01	Multiple Choice
39	dat540.22h.01	Multiple Choice

40	dat540.22h.01	Multiple Choice

1 dat540.22h.04

We want to sort the columns of a dataframe df based on their names, how should we do it?
Select one alternative:

- df.sort_index()
- df.sort()
- df.sort_index(axis=1)
- df.sort(axis='columns')

2 dat540.22h.04

Which one is FALSE about Gradient Descent for an ANN?
Select one alternative:

- If Learning Rate is very small, it will take a lot of time to optimize the cost function
- If Learning Rate is very large, Gradient Descent may never converge
- Learning Rate is calculated using Loss Function, Derivative and Step Size
- Learning Rate is a hyper-parameter in Gradient Descent algorithm

3 dat540.22h.04

Which of the following algorithms is best suited to predict whether it will rain today?
Select one alternative:

- Logistic Regression
- Linear Regression
- All of the options
- K-Nearest Neighbor

4 dat540.22h.04

Observe the following code and identify what will be the outcome?

```
arr = [['DAT540', '450', 'Python'], [120, 450, 540]]
arr[0][1][::2]
```

Select one alternative:

- ‘DT4’
- ‘45’
- None of the options
- ‘40’

5 dat540.22h.04

We want to display a labeled plot and we have the code, but the order of execution is wrong. Which one is the correct order of execution to do this?

1. plt.title('DAT540')
2. data.plot()
3. plt.show()
4. import matplotlib.pyplot as plt

Select one alternative:

- 2-1-3-4
- 4-1-2-3
- 3-4-2-1
- 4-2-1-3

6 dat540.22h.04

Observe the following code and identify what will be the outcome?

```
def total(initial = 5, *num, **key):
    count = initial
    for n in num:
        count+=n
    for k in key:
        count+=key[k]
    return count


print(total(150,5,7, clouds=10, stars=35))
```

Select one alternative:

- 107
- 207
- 62
- 325

7 dat540.22h.04

Observe the following code and identify what will be the outcome?

```
import math

def main():
    math.cos(math.pi)

print(main())
```

Select one alternative:

- Error
- -1.0
- None
- None of the options

8 dat540.22h.04

Observe the following code and identify what will be the outcome?

```
a = 540
def f():
    global a
    print('a: ', a)
    a=240
    print('new a: ', a)

f()

print('value of a: ', a)
```

a). a: 540
new a: 240
value of a: 240

b). a: 540
new a: 540
value of a: 540

c). a: 540
new a: 240
value of a: 540

d). a: 540
new a: 540
value of a: 240

Select one alternative:

- a
- b
- c
- d

9 dat540.22h.04

We have two lists

```
a = [1,2,3]
```

```
b = [4,5,6]
```

How can we produce this list from these two lists:

```
a = [1,2,3,4,5,6]
```

Select one alternative:

- a.append(b)
- a.extend(b)
- a = a + b
- a.concat(b)

10 dat540.22h.03

What can be caused by the presence of outliers in the dataset?
Select one alternative:

- False mean value
- Overfitting
- Skewed standard deviation
- All of the options

11 dat540.22h.03

We have a dataframe named 'purchase' as follows:

	name	itemID	value
0	John	AB	10
1	Alice	AY	35
2	Ashley	AC	56
3	Bob	AY	90
4	Bob	A5	27
5	Bob	A3	48
6	John	A2	55
7	Ashley	A4	61
8	Alice	A6	70
9	Bob	AX	84

Which of the following implementation would add a new column named 'rank', to the existing dataframe where the value of the column 'rank' would be rank of each 'itemID' **within** each 'name' based on 'value' in decending order? The output dataframe would look like:

	name	itemID	value	rank
0	John	AB	10	2.0
1	Alice	AY	35	2.0
2	Ashley	AC	56	2.0
3	Bob	AY	90	1.0
4	Bob	A5	27	4.0
5	Bob	A3	48	3.0
6	John	A2	55	1.0
7	Ashley	A4	61	1.0
8	Alice	A6	70	1.0
9	Bob	AX	84	2.0

Select one alternative:

- purchase["rank"] = purchase.groupby("name")["value"].rank(ascending=False)
- purchase["rank"] = purchase.groupby("name")["value"].rank("first")
- purchase["rank"] = purchase.groupby("name")["value"].rank()
- purchase["rank"] = purchase.groupby("name")["value"].rank("dense")

12 dat540.22h.03

Which of the following is a true statement?

Select one alternative:

- Pandas series cannot have same value as index for multiple rows
- By default, pandas series automatically assigns column indices from 0 to N (where N is the length of the data)
- A pandas series can be conceptualized as a fixed length ordered dictionary
- A pandas series is a list-like object containing a sequence of values

13 dat540.22h.03

Which of the following numpy function, we can use to return the list of indexes in order of the sorted values?

Select one alternative:

- np.lexsort()
- np.argsort()
- np.idxsort()
- sorted()

14 dat540.22h.03

Which of the following is not a Numpy Universal function (ufunc) example?

```
x = np.random.randn(100)
y = np.random.randn(100)
```

Select one alternative:

- np.accumulate(y)
- np.modf(y)
- np.maximum(x,y)
- np.setdiff1d(x,y)

15 dat540.22h.03

What does matplotlib.pyplot.add_subplot(3, 2, 1) do?
Select one alternative:

- Adds a subplot at first position of a 3 x 2 plotting space
- Adds a subplot at first position of a 2 x 3 plotting space
- Adds a subplot at third position of a 2 x 1 plotting space
- Adds a subplot at third position of a 1 x 2 plotting space

16 dat540.22h.03

We have a dataframe named ‘temparature’ which shows actual temperature and how it feels like for different cities of Norway.

	Temparature	
	actual(°C)	feels like(°C)
Stavanger	13	10
Oslo	10	7
Bergen	11	7

Which of the following implementation would give us the reshaped dataframe keeping ‘Temparature’ as the column and grouping ‘actual ()’ and ‘feels like ()’ with the respective cities? The resultant dataframe would look like this:

		Temparature
Stavanger	actual(°C)	13
	feels like(°C)	10
Oslo	actual(°C)	10
	feels like(°C)	7
Bergen	actual(°C)	11
	feels like(°C)	7

Select one alternative:

- temparature.groupby(axis=0, level=0)
- temparature.stack(level=-1)
- temparature.reshape(axis=0, level=0)
- temperature.swaplevel('actual(°C)', 'feels like(°C)')

17 dat540.22h.03

Which one is the built-in module in Python for serialization?
Select one alternative:

- msgpack
- feather
- pickle
- HDF5

18 dat540.22h.03

We have a dataframe as follows:

```
rand_arr = np.random.randint(0, 100, 40).reshape(8,5)
data = pd.DataFrame(rand_arr, columns=['A', 'B', 'C', 'D', 'E'] )
```

Which of the following implementation would give us column B, C, D where the value of column C is greater than 33?
Select one alternative:

- data[data.C > 33].iloc[:, 1:4]
- data[data.C > 33].loc[:, ‘B’:‘E’]
- data[data.C > 33].iloc[:, ‘B’:‘E’]
- data[data.C > 33].loc[:, 1:4]

19 dat540.22h.03

Which one of the following representation is ‘bare slice assignment’ on the following numpy array?

```
arr = np.arange(10)
```

Select one alternative:

- arr[0:-1] = 1
- arr[1:] = 1
- arr[:] = 1
- arr[:-1] = 1

20 dat540.22h.02

Given the following arrays, what would be the outputs of x*y and x*z.

```
x = np.array([5.9, 6.2, 55.2])
y = np.array([10.0, 10.0, 10.0])
z = 10.0
```

Select one alternative:

- [59.0, 62.0, 552.0] and 59.0
- [59.0, 62.0, 552.0] and [59.0]
- [59.0, 62.0, 552.0] and [59.0, 62.0, 552.0]
- None of the options

21 dat540.22h.02

Given the following code, analyze to get the output of variable x.

```
x = [1, 2, 3, 4]
l = lambda x: (x + 5)

for i in range(len(x)):
  l(x[i])
```

Select one alternative:

- [1, 2, 3, 4]
- [6, 7, 8, 9]
- [5, 7, 8, 9]
- None of these

22 dat540.22h.02

What in general terms you understand overfitting as?
Select one alternative:

- Performs well in test data, but not in training data.
- Performs well in training data, but not in test data.
- Performs well in both train and test data.
- None of the options

23 dat540.22h.02

Given the arrayValues, what would be the output of x?

```
arrayValues = np.array([8, 7, 6, 5, 4, 3, 2, 1])

x = np.where(arrayValues % 2 == 1)
```

Select one alternative:

- 7, 5, 3, 1
- 8, 6, 4, 2
- 0, 2, 4, 6
- 1, 3, 5, 7

24 dat540.22h.02

Given the dataframe, what are the class outputs of df.iloc[0] and df.loc[0].

```
df = pd.DataFrame({'class': ['Cat', 'Dog', 'Bus']},
index=[7, 0, 1])
```

Select one alternative:

- Dog and Dog
- Cat and Cat
- Cat and Dog
- Dog and Cat

25 dat540.22h.02

Given the following numpy array, where arange would generate the number between 0 to 5, what would be the output of the indexing.

```
X = np.arange(6).reshape((2, 3))
X[-2, [2, 1]]
```

Select one alternative:

- [2, 1]
- [4, 5]
- [0, 1]
- [3, 4]

26 dat540.22h.02

What would be the output of the sales_dict,

```
sales_dict = {
   'Oslo': 450,
   'Stavanger': 5855,
   'Bergen': 585,
   'Oslo': 858,
   'Stavanger': 857
}
sales_dict['Oslo'] += 5

sales_dict
```

Select one alternative:

- {'Oslo': 455, 'Stavanger': 5855, 'Bergen': 585}
- {'Oslo': 1358, 'Stavanger': 6712, 'Bergen': 585}
- {'Oslo': 863, 'Stavanger': 857, 'Bergen': 585}
- Code outputs error.

27 dat540.22h.02

Given the following data frame sales,

```
sales = pd.DataFrame({'datetime': ['20220225', '20220226', '20220226',
'20220225'],'sales': [150, 250, 350, 450]})
```

How would you print the average sales of the datetime.
Select one alternative:

- pd.pivot_table(sales, index=["datetime"])
- sales.groupby(['datetime']).mean()
- pd.pivot_table(sales, index=["datetime"], aggfunc=np.sum)
- All of the options

28 dat540.22h.02

What is the correct output of the variables, c and d.

```
def divideby(a = 6):
   b = float(48)
   return (b / a)

value = '2'
c = divideby(float(value))
d = divideby()
```

Select one alternative:

- 8.0 and 8.0
- 24.0 and 8.0
- 8.0 and 24.0
- TypeError (Unsupported operand types for / : ‘float and str’)

29

dat540.22h.02

Given a dataframe

```
figframe = pd.DataFrame({
   "key" : ['A', 'B', 'C', 'A', 'B', 'C'],
   "data" : [1, 1, 2, 2, 3, 3]
})
```

What would be the output if you groupby “key” and apply sum to it.

```
figframe.groupby("key").apply(sum)
```

a).

	data
data	
1	2
2	4
3	6

b).

	key	data
key		
A	AA	3
B	BB	4
C	CC	5

c).

	key	data
data		
1	AB	2
2	CA	4
3	BC	6

d).

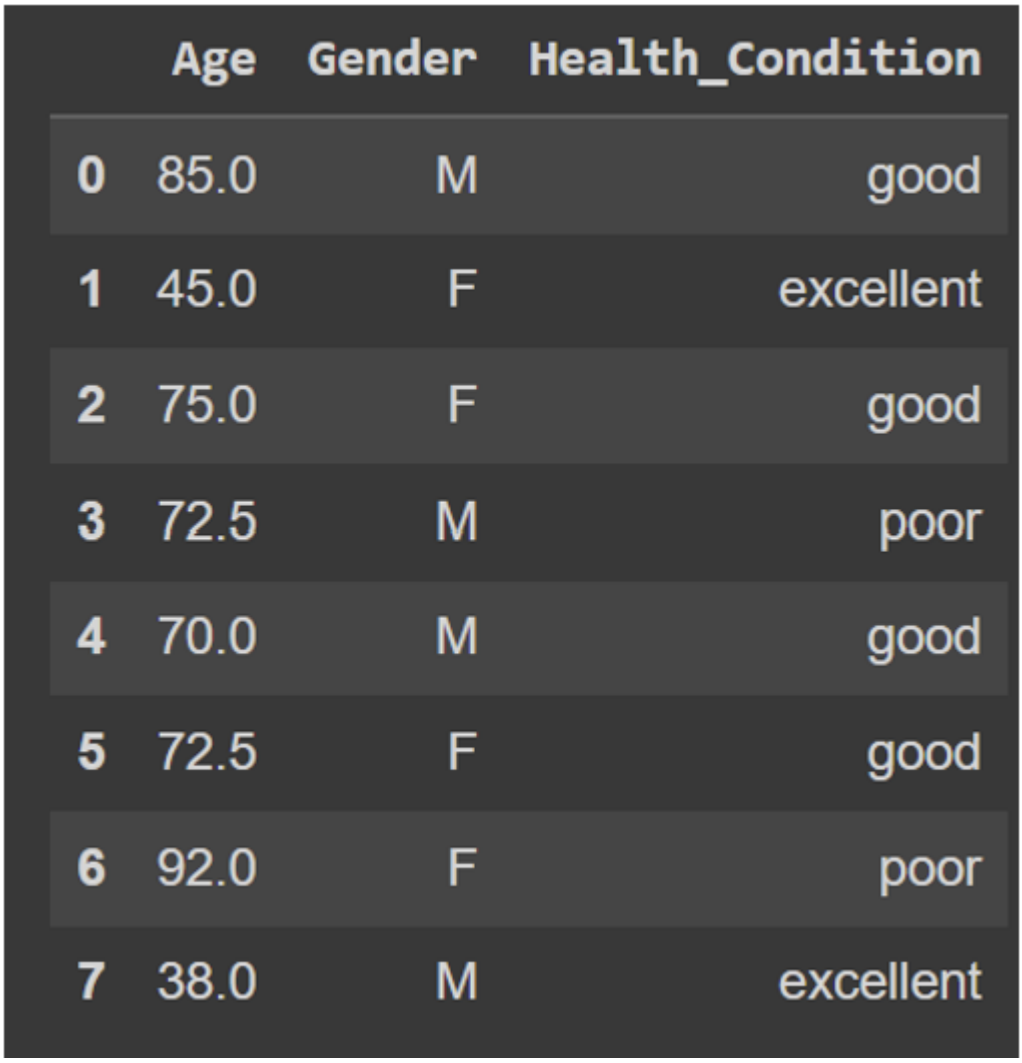
	data
key	
A	3
B	4
C	5

Select one alternative:

- a
- b
- c
- d

30 dat540.22h.01

For a given data frame (df), Which of the following code will convert the column to pandas series?



	Age	Gender	Health_Condition
0	85.0	M	good
1	45.0	F	excellent
2	75.0	F	good
3	72.5	M	poor
4	70.0	M	good
5	72.5	F	good
6	92.0	F	poor
7	38.0	M	excellent

Select one alternative:

- df[['Health_Condition']]
- df['Health_Condition']
- Both of the options
- None of the options

31 dat540.22h.01

What would the print statement show in this case?

```
import numpy as np
a = np.zeros((2,3))
b = np.ones((2,3))
c = a + b
c = (c**2).T
print (c)
```

Select one alternative:

- [[1. 1. 1.] [1. 1. 1.]]
- [[2. 2. 2.] [2. 2. 2.]]
- [[1. 1.] [1. 1.] [1. 1.]]
- [[2. 2.] [2. 2.] [2. 2.]]

32 dat540.22h.01

What is the output of the size method applied to a numpy matrix A?
A = [5,6,1,8,6]

Select one alternative:

- 5
- (5,)
- (5,1)
- All of the options

33 dat540.22h.01

What would be the value of b in the following code after execution?

```
a = np.array([[3,5,7,9], [13,78,2, 4]])
b = (a%3==1)
```

Select one alternative:

- array([[False, False, True, False], [True, False, False, True]])
- array([[True, True, True, True], [True, False, False, False]])
- array([[True, False, False, True], [False, True, False, False]])
- None of the options

34 dat540.22h.01

Which of the following scenarios presents over-fitting?
Select one alternative:

- Low Variance
- High Variance
- None of the options
- All of the options

35 dat540.22h.01

Which of the following statement will return True if func is defined as follow

```
def func(input):
    if isinstance(input,str):
        pass
    else:
        return ('Enter a valid string')
    for i in range (0, int(len(input)/2)):
        if input[i].lower() != input[-i-1].lower():
            return ('Not Palindrome')
    return True
```

Select one alternative:

- func(0)
- func("text")
- func("Redder")
- None of the options

36 dat540.22h.01

Which of the following exception-handling code will provide a correct print message (instead of throwing an error).

```
import pandas as pd
import numpy as np
a = pandas.Series(np.array([5,74]))
```

a).

```
try:
  import pandas as pd
  import numpy as np
  a = pandas.Series(np.array([5,74]))
except ImportError:
  print("ImportError exists")
```

b).

```
try:
  import pandas as pd
  import numpy as np
  a = pandas.Series(np.array([5,74]))
except NameError:
  print("NameError exists")
```

c).

```
try:
  import pandas as pd
  import numpy as np
  a = pandas.Series(np.array([5,74]))
except IOError:
  print("IOError exists")
```

d).

```
try:
  import pandas as pd
  import numpy as np
  a = pandas.Series(np.array([5,74]))
except SyntaxError:
  print("SyntaxError exists")
```

Select one alternative:

- a
- b
- c
- d

37 dat540.22h.01

What would be the output of the following code:

```
try:
  a=10
  b=0
  div = a/b
  print(div)
except:
  div = b/a
  print(div)
```

Select one alternative:

- ZeroDivision Error
- 0
- Both ZeroDivision Error and 0
- None of these

38 dat540.22h.01

For two data frames A & B, what would be the result of cross merge (pd.merge(A,B,how="cross")).

A

	A	B
0	5	11
1	16	22

B

	C	D
0	15	1
1	18	33

a).

	A	B
0	15	1
1	18	33

c).

	A	B
0	5	11
1	16	22
2	15	1
3	18	33

b).

	A	B	C	D
0	5	11	15	1
1	16	22	18	33

d).

	A_x	B_x	A_y	B_y
0	5	11	15	1
1	5	11	18	33
2	16	22	15	1
3	16	22	18	33

Select one alternative:

- a
- b
- c
- d

39 dat540.22h.01

For the following code, which output is going to be displayed

```
try:
   lis = 3*[1]+1*[3]
   x = lis[10]
   print('A!')
except IndexError:
   print('B! ')
else:
   print('C!')
finally:
   print('D!')
```

Select one alternative:

- A! & D!
- B! & D!
- C! & D!
- B!

40 dat540.22h.01

Which of the following *print* statement will print the correct value from variable “accuracy” up to 2 decimal places if the accuracy value is 0.98?

Select one alternative:

- print(f"Your model has accuracy value: {accuracy}")
- print("Your model has accuracy value:", accuracy)
- print("Your model has accuracy value: {:.2f}".format(accuracy))
- All of the options