

**Disclaimer:** This is the first semester in which Professor Cannon has ever held a final exam for 1002, the result of which is that I and no one other than him has any idea what the content of the final will be like. As a result I cannot say with certainty that the questions you see on this document are going to be like those on the final, what I can guarantee is that I wrote these questions after a deep dive into the lectures from the context, meaning the content presented is stuff you should be aware of whether it is programming related or purely mathematical. I also plan to make this a full length practice exam so there will be 25 multiple choice questions, you will have the class time (technically slightly less) in order to complete the exam so please simulate the actual testing environment when you complete this practice. As always, Good luck!

COMS 1002 Computing in Context Practice Final Exam

Name: \_\_\_\_\_

Please Use **CAPITAL LETTERS** to fill in the boxes: A B C D E

1.	2.	3.	4.	5.
6.	7.	8.	9.	10.
11.	12.	13.	14.	15.
16.	17.	18.	19.	20.
21.	22.	23.	24.	25.

1. Consider the popular data science module 'Pandas'. How can we enable the use of pandas in our python programs?

- a. `import module pandas`
- b. `from pandas import pd`
- c. `import numpy as np`
- d. `import pandas_data`
- e. `import pandas`

2. Consider the head of the dataframe *df* below, how can we change the index to be the second column from the left?

	name	email	address
0	Donald	donald@magic.kingdom	NaN
1	Goofy	goofy@magic.kingdom	NaN
2	Ariel	ariel@magic.kingdom	NaN
3	Minnie	minnie@magic.kingdom	NaN
4	Mickey	mickey@magic.kingdom	NaN
5	Ursula	ursula@magic.kingdom	NaN
6	Captain Hook	NaN	The seven seas

- a. `df.set_index("email")`
- b. `df.index("email")`
- c. `df.set_index(email.toString())`
- d. `df = df.index("email")`
- e. `df = df.set_index("email")`

3. We have two methods meant for locating information within a dataframe: what are they and what is the primary difference between them?

- a. `.locate()` and `.ilocate()` the former uses labels to find the information and the latter uses integer based indexing
- b. `.loc()` and `.iloc()` the former uses labels to find the information and the latter uses integer based indexing
- c. `.loc[]` and `.iloc[]` the former uses labels to find the information and the latter uses integer based indexing
- d. There is only one way to locate values in a Pandas dataframe
- e. None of the above are correct

4. Consider the following code fragment, what is the output when the code is run in a proper development environment?

```
import numpy as np

a = np.array([1,2,3])
b = np.array([[4,5,6],[7,8,9]])
c = a*b
c = c.reshape(3,2)

print(c)
```

- a.  $\begin{bmatrix} 4 & 10 & 18 \\ 7 & 16 & 27 \end{bmatrix}$
- b.  $\begin{bmatrix} 4 & 10 & 18 \\ 7 & 16 & 27 \end{bmatrix}$
- c. This code will result in an error
- d.  $\begin{bmatrix} 4 & 10 \\ 18 & 7 \\ 16 & 27 \end{bmatrix}$
- e.  $\begin{bmatrix} 4 & 10 \\ 18 & 7 \\ 16 & 27 \end{bmatrix}$

5. Recall that we can define functions to help us determine the probability of certain events. In particular we can consider the fact that a standard coin flip can be modeled with a Bernoulli with a probability of success of 0.5. Now consider a biased coin where the probability of getting heads is 0.25, what is the probability that I flip 10 heads in a row?

- a.  $(0.5)^{10}$
- b.  $(0.75)^{10}$
- c.  $(0.25)^{10}$
- d.  $(0.25)^{10} / 2^{10}$
- e.  $(0.75)^{10} / 2^{10}$

6. Recall your discussion on argument and parameter types in Python. Using that knowledge consider the following code segment and determine what the output is:

```
def foo(a, b, c, *args, d = 0, e = 100):  
    sum = 0  
    for value in args:  
        sum += (value * d)  
    sum -= e  
    sum *= b  
    sum /= a  
    return sum
```

foo(2, 5, 8, 9, 2, 150)

- a. -250.0
- b. 430.0
- c. -375.0
- d. 555.0
- e. The code produces an error

7. Recall from a previous economics lab that we wrote some functions in order to calculate the stock price  $S$  after  $T$  days using the following formula:

$$S = S_0 Y_1 Y_2 \dots Y_T$$

Specifically recall that you may have used `np.product` to compute the product of a list of values, which of the following is a proper alternative for that specific part of the code.

- a.  

```
prod = 1
for val in some_list:
    prod += val
```
- b.  

```
prod = 1
for val in some_list:
    prod *= val
```
- c.  

```
prod = 1
for val in some_list:
    prod **= val
```
- d. `prod = np.sum(some_list)`
- e. None of the above are algorithmic substitutes

8. When discussing algorithms we have a systematic approach in order to describe how efficient they are. Recall *\*Linear\** Search from lecture what is the big- $O$  of that algorithm.

- a.  $O(n)$
- b.  $O(1)$
- c.  $O(n^3)$
- d.  $O(\log n)$
- e.  $O(n^2)$

9. What is the value of `y` after the following code executes:

```
x = "Columbia University in the City of New York"
y = x[4:12:3]
```

- a. "man"
- b. "mane"
- c. "uiU"
- d. "m v"
- e. There is an error in this code

10. Consider some standard pandas data frame `df` and determine what is printed to the console:

```
type(df.iloc[0])
```

- a. <class 'pandas.core.series.Series'>
- b. <class 'pandas.core.frame.DataFrame'>
- c. <class 'numpy.int64'>
- d. <class 'str'>
- e. Cannot be determined

11. Consider the partial function definition below and answer the following question:

```
def foo(w,x,*y,**z):
```

What is the data type of `z` in the above function?

- a. tuple
- b. string
- c. dictionary
- d. set
- e. Cannot be determined

12. Consider the function below:

```
def bar(x, y):  
    if y == 1:  
        return 1  
    return 1 + x(x, y-1)
```

What is the result of bar(bar, 5) ?

- a. 5
- b. 0
- c. 1
- d. Cannot be Determined
- e. There is an error in the code

13. What is stored in d after the following code block executes?

```
import numpy as np  
  
a = np.ones((3,3))  
b = np.array(range(9))  
b.shape = (3,3)  
c = 2+(a*b)  
d = 2*(c+b)
```

- a. 

```
[[ 0. 10. 32.]  
 [ 66. 112. 170.]  
 [240. 322. 416.]]
```
- b. 

```
[[ 2.  5.  8.]  
 [11. 14. 17.]  
 [20. 23. 26.]]
```
- c. 

```
[[ 2.  3.  4.]  
 [ 5.  6.  7.]  
 [ 8.  9. 10.]]
```
- d. 

```
[[ 4.  8. 12.]  
 [16. 20. 24.]  
 [28. 32. 36.]]
```
- e. Trick question; you can't multiply arrays

14. When performing Cross-Validation, specifically a k-fold, what portion of the data is used as training data within an iteration ?

- a. k
- b. 0
- c.  $k - 1$
- d.  $k // 2$
- e. We do not use training data when performing cross validation

15. When performing Cross-Validation, specifically a k-fold, what portion of the data is used as testing data within an iteration ?

- a. 1
- b. 0
- c.  $k - 1$
- d.  $k - 1 // 2$
- e. We do not use testing data when performing cross validation

16. During lecture on 11/30/2023 you had a brief introduction into machine learning, which of the following was not a classifier type mentioned during lecture.

- a. K-Nearest Neighbors
- b. Cross validation
- c. Support Vector Machines
- d. Random Forest Classifiers
- e. All of these are classifiers



17. What is the difference between supervised and unsupervised learning?

- a. Supervised learning involves labeled data, while unsupervised learning deals with unlabeled data.
- b. In supervised learning, the algorithm is trained on input-output pairs, while unsupervised learning finds patterns in data without labeled outputs.
- c. Supervised learning is used for classification and regression tasks, whereas unsupervised learning is employed for clustering and dimensionality reduction.
- d. In supervised learning, the model learns from a teacher (labeled data), while in unsupervised learning, the model explores the data structure without explicit guidance.
- e. Supervised learning requires a target variable for training, while unsupervised learning discovers inherent patterns or structures in the data without predefined target variables.

18. Recall any implementation you have seen for the binomial random variable whether this be from lecture or from the homeworks. Assume it is called `binom` and takes two arguments `n` and `p` such that the following code properly executes.

```
n = 100, p = 0.25
x = binom(n, p)
```

Say we wanted to validate that the value `x` ends up being what we expect it to be, how would we do this and what is the expected value we are looking for?

- a. We can print out the value of `x` and compare it to the expected value `p`
- b. We can use monte carlo simulations along with the law of large numbers and print the result, comparing it to the expected value of `np`
- c. We can use a for loop and add up `N` results of `binom` and print out the result and compare it to the expected value `Np`
- d. There is no way to know what the expected value is, for it is truly random.
- e. None of the Above

19. What values are possible from a Bernoulli random variable with a success probability  $p$ ? You can think of a python implementation but it is not necessary.

- a.  $[0, 1]$
- b.  $[\text{True}, \text{False}]$
- c.  $[0, p]$
- d.  $[p // 2]$
- e. None of the above

20. Consider the following two collections of data points, each representing an axis.

```
x_data = [1, 3, 5, 9, 12]
y_data = [1, 9, 25, 81, 144]
```

If we were to plot this with matplotlib, what we we expect the shape of the line through these points to look like.

- a. Linear
- b. Cubic
- c. Quadratic
- d. Quartic
- e. No relationship can be predetermined

21. Assume we purchased a European put option and our goal was to ensure a profit, which of the following inequalities would indicate that we made a profit.

- a.  $K \leq S_0(1+r)^T$
- b.  $K < S_0(1+r)^T$
- c.  $K < S_T$
- d.  $K = S_T$
- e.  $K > S_T$

22. When we discussed presidential elections we discussed the probability of there being a tie between two candidates, when you simulated this on the homework you divided the total ties by the trials then multiplied by some quantity let's call this quantity  $M$

Now assume that in the present year 2023 we gained Canada, Puerto Rico, and Greenland as states (sorry DC) and assume the total amount of votes in the EC stay the same, what will happen to  $M$  and how will this affect the earlier calculations?

- a. The calculation would see a decrease in the amount since  $M$  is inversely proportional to the number of states.
- b. The calculation would stay the same since  $M$  increases.
- c. The calculation would stay the same since  $M$  does not depend on the amount of states.
- d. The calculation would increase since  $M$  increases.
- e. I have no idea what you are referring to

23. What is the difference between a function and a method in Python?

- a. Methods are functions that operate on objects and are equipped with the implicit parameter
- b. Functions and Methods are 100% equivalent
- c. Methods are functions that take only a single parameter
- d. Functions are defined using `def` while methods are not
- e. None of the Above.

24. What would be the value of the variable `list2` after the execution of the following code:

```
list1 = [1,2,3]
list2 = sorted(list1)
list1[1] = 5
```

- a. `[1,3,2]`
- b. `[1,2,3]`
- c. `[3,2,1]`
- d. `list2` is no longer assigned a value.
- e. There is an error in the code above.

25. What is printed to the terminal after the following code is executed?

```
x = 2023
```

```
y = 334
```

```
while x > y:
```

```
    x -= y
```

```
    y += 2
```

```
print("This is the grade on your test: ", ((y - x - 10) * 100))
```

- a. This is the grade on your test: -989
- b. This is the grade on your test: 100
- c. This is the grade on your test: -31956
- d. There is an error in the code segment
- e. None of the above

### Answer Key:

1. E: `import pandas`
2. E: `df = df.set_index("email")`
3. C: `.loc[]` and `.iloc[]` the former uses labels to find the information and the latter uses integer based indexing
4. E: 

```
[[ 4 10]
 [18  7]
 [16 27]]
```
5. C:  $0.25^{10}$
6. A:  $-250.0$
7. B: `prod = 1 for val in some_list: prod *= val`
8. A:  $O(n)$
9. A: "man"
10. A: `<class 'pandas.core.series.Series'>`
11. C: dictionary
12. A: 5
13. D: 

```
[[ 4.  8. 12.]
 [16. 20. 24.]
 [28. 32. 36.]]
```
14. C:  $K - 1$
15. A: 1
16. B: Cross validation
17. A: Supervised learning involves labeled data, while unsupervised learning deals with unlabeled data.
18. B: We can use monte carlo simulations along with the law of large numbers and print the result, comparing it to the expected value of  $np$
19. A:  $[0, 1]$
20. C: Quadratic
21. B:  $K < S_0(1+r)^T$
22. D: The calculation would increase since  $M$  increases.
23. A: Methods are functions that operate on objects and are equipped with the implicit parameter
24. B:  $[1,2,3]$
25. B: This is the grade on your test: 100

A Final Note to Students: If you manage to get this far, thank you and good luck on your exam!