

CS 320 Exam 3 (15%) - Spring 2022

Instructor: Tyler Caraza-Harter

First/Given Name: _____. Last/Surname: _____

Net ID: _____@wisc.edu

Fill in these fields (left to right) on the scantron form (use pencil):

1. LAST NAME (surname) and FIRST NAME (given name), fill in bubbles
2. IDENTIFICATION NUMBER is your Campus ID number, fill in bubbles
3. Under A of SPECIAL CODES, write your lecture number, fill in bubbles. 1=8:50am, 2=11am
4. Under B of SPECIAL CODES, tell us about the nearest person (if any) to your left. 0=no person to the left in your row, 1=somebody you do not know is there, 2=somebody you do know is there.
5. Under C of SPECIAL CODES, do the same as B, but for the person to your right
6. **Under D of SPECIAL CODES, write 1 and fill in bubble 1.** This is very important!

Make sure you fill all the special codes above accurately in order to get graded.

You have 2 hours to take the exam. Use a #2 pencil to mark all answers. When you're done, please hand in these sheets in addition to your filled-in scantron. You may not sit adjacent to your friends or other people you know in the class (having only one empty seat is considered "adjacent"). You may only reference your notesheet. You may not use books, your neighbors, calculators, or other electronic devices on this exam. Turn off and put away portable electronics now. If multiple answers to a question are correct, choose the best answer.

(Blank Page for You to Do Scratch Work)

Q1. What are the bounds on possible values for `s`?

```
lreg = LinearRegression()
lreg.fit(train[xcols], train[ycol])
s = lreg.score(test[xcols], test[ycol])
```

- (A) 0 to 1 (B) -1 to 1 (C) -infity to 1 (D) -infinity to infinity (E) 0 to infinity

Q2. What is the complexity of the following code, if `N` is the length of the list `L`? Choose the best answer.

```
for val in L:
    threshold = max(L)
    if val >= 4 * threshold:
        print("hi!")
```

- (A) $O(1)$ (B) $O(N)$ (C) $O(N^2)$ (D) $O(N^2 + 1)$ (E) $O(N^3)$

Q3. If we can successfully call `obj.fit(df)`, what type could `obj` be?

- (A) LinearRegression only
(B) LinearRegression or LogisticRegression
(C) KMeans only
(D) KMeans or AgglomerativeClustering

Q4. If you want to produce a dendrogram, what should you use?

- (A) LinearRegression (B) LogisticRegression (C) KMeans (D) AgglomerativeClustering (E) PCA

Q5. In the following code analyzing a contingency table from an A/B test, `pvalue` is 0.055. The threshold for significance is 0.05. Do we have statistically significant evidence that B has a different click-through-rate than A?

```
from scipy import stats

df = pd.DataFrame({
    "click": {"A": ????, "B": ????,},
    "no-click": {"A": ????, "B": ????,},
})

_, pvalue = stats.fisher_exact(df)
```

- (A) yes (B) no

Q6. True or False? The B class is a child of the A class; both have an `__init__` method. Both `__init__` methods are guaranteed to run when a new instance of B is created, regardless of the code in B's `__init__` method.

- (A) True (B) False

Q7. What call makes predictions using a computation similar to `x @ c`?

- (A) `LinearRegression.predict` (B) `LogisticRegression.predict`
(C) `LinearRegression.predict_proba` (D) `LogisticRegression.predict_proba`

Q8. A polygon with alpha 1 is hidden beneath a diamond. If you want to see the polygon, what could you do?

- (A) decrease zorder of the polygon
(B) increase zorder of the diamond
(C) increase alpha of the diamond
(D) increase alpha of the polygon
(E) decrease zorder of the diamond

Q9. Every non-leaf node in a binary search tree has at most two children.

- (A) True (B) False

Q10. The shape of `A` is (9, 6), the shape of `B` is (6, 8), and the shape of `C` is (8, 5). What is the shape of `A@B@C`?

- (A) (6, 8) (B) (8, 5) (C) (9, 5) (D) (9, 6)

Q11. What is an advantage that a typical GPU has over a typical CPU?

- (A) more cores (B) more flexible cores (C) cores with a faster clockrate (D) all of the above

Q12. What is `x`?

```
class C:
    def __init__(self, vals):
        self.vals = vals

    def __len__(self):
        return 4

    def __getitem__(self, lookup):
        return 2

obj = C([1, 3, 5])
x = len(obj.vals) # careful!
```

- (A) 1 (B) 2 (C) 3 (D) 4 (E) 5

Q13. How many oranges are classified as apples, according to the following confusion matrix?

	apples	oranges	bananas
apples	35	5	97
oranges	43	67	70
bananas	16	80	69

(A) 5 (B) 38 (C) 43 (D) 48

Q14. What scikit-learn transformer could have potentially produced the following output features?

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

(A) PolynomialFeatures (B) OneHotEncoder (C) StandardScaler

Q15. What is a valid simplification of the following, assuming the code runs without error?

```
X @ np.linalg.solve(X, y)
```

(A) \bar{X} (B) \bar{y} (C) \bar{c} (D) $X @ y$

Q16. If a Flask app has the following handlers, what does it print when a user visits the home page in a browser?

```
@app.route("/")
def home():
    print("X")
    return '<html><body></body></html>'
@app.route("/example.svg")
def handler1():
    print("Y")
    return "TODO"
@app.route("/plot.svg")
def handler2():
    print("Z")
    return "TODO"
```

(A) X only (B) X and Y (C) X and Z (D) X, Y, and Z

Q17. `df` has 15 columns and 107 rows. After running the following, how many columns will `p.components_` have?

```
p = PCA(4)
p.fit(df)
```

(A) 4 (B) 15 (C) 60 (D) 107

Q18. In complexity analysis, which statement about steps is true?

- (A) all steps must take the same amount of time to execute
- (B) a step's execution time never depends in any way on the input
- (C) a piece of code containing a loop always consists of multiple steps
- (D) a step has a bounded execution time that doesn't keep growing with input size

Q19. You want to create a model to predict a number. What kind of machine learning task is this?

- (A) regression
- (B) classification
- (C) clustering
- (D) decomposition

Q20. What is printed?

```
items = []  
for item in [0.5, -1, 5, 0, 2]:  
    heapq.heappush(items, item)  
print(heapq.heappop(items))
```

- (A) 2
- (B) -1
- (C) 0
- (D) 0.5

Q21. If $A = \text{np.array}([[2, 6], [3, 1]])$ and $b = \text{np.array}([[5], [4]])$, what is $A*b$?

- (A) $\begin{bmatrix} [10, 24] \\ [15, 4] \end{bmatrix}$
- (B) $\begin{bmatrix} [10, 30] \\ [12, 4] \end{bmatrix}$
- (C) $\begin{bmatrix} [18, 54] \\ [27, 9] \end{bmatrix}$
- (D) $\begin{bmatrix} [34] \\ [19] \end{bmatrix}$

Q22. If a BST is constructed using the algorithm we learned in class, and the insert order is [7, 8, 5, 9], where will 9 be?

- (A) root.left.left
- (B) root.left.right
- (C) root.right.left
- (D) root.right.right

Q23. True or False. A generator object returned by a generator function supports indexing, slicing, and looping, just like a list.

- (A) True
- (B) False

Q24. Is the center of the text created by the following code within the region bounded by the ax subplot?

```
ax.text(0.1, 0.9, "hi", transform=ax.transAxes, size=20, color="red",  
        verticalalignment="center", horizontalalignment="center")
```

- (A) definitely not
- (B) definitely
- (C) it depends on the xlim and ylim of the ax region

Q25. Assume the following code runs successfully on some unknown string. What letter(s) could possibly be printed?

```
matches = re.findall(r"(A)B([CD])", some_string)
print(matches[-1][0])
```

(A) A only (B) B only (C) C only (D) C or D (E) A, C, or D

Q26. What does `nums` contain after the following runs?

```
nums = []
def count(z):
    if z < 4:
        count(z+1)
    nums.append(z)
count(1)
print(nums)
```

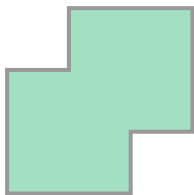
(A) [4, 3, 2] (B) [] (C) [1, 2, 3, 4] (D) [2, 3, 4] (E) [4, 3, 2, 1]

Q27. Does the regular expression `r"\d.*?[A-Z]"` match anything in the string `"cB1adb."`?

(A) yes (B) no

Q28. What could be added to the following code to produce the below shape?

```
from shapely.geometry import box, Point
x = box(0, 0, 2, 2)
y = box(1, 1, 3, 3)
```



(A) `x.union(y)` (B) `x.intersection(y)` (C) `x.difference(y)` (D) `y.difference(x)`

Q29. What is the recall for oranges, given the following confusion matrix?

	apples	oranges	bananas
apples	0	1	5
oranges	5	3	2
bananas	0	3	1

(A) 0.3 (B) 0.4 (C) 3 (D) 10

Q30. You want to extract a subset of a DataFrame's columns to produce a new DataFrame, to which you'll fit KMeans. What should go between the brackets?

- (A) list of strings
- (B) a make_column_transformer
- (C) a string, equal to the name of the label column
- (D) a string, equal to the name of a feature column