CS 320 Final (20%) - Fall 2021

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 2 and fill in bubble 2**. 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. Please place your student ID face up on your desk. Turn off and put away portable electronics now.

(Blank Page for You to Do Scratch Work)

Q1. For what is OneHotEncoder most useful?

(A) numeric labels (B) numeric features (C) categorical labels (D) categorical features

Q2. If you want an animation to be smoother without changing the total time, what should you do?

(A) decrease interval and decrease frames
(B) decrease interval and increase frames
(C) increase interval and decrease frames
(D) increase interval and increase frames

Q3. Which of the following must be given "correct" categories when `fit` is called?

(A) AgglomerativeClustering (B) KMeans (C) LogisticRegression (D) LinearRegression

Q4. What is the recall for bananas, given the following confusion matrix?

	apples	oranges	bananas
apples	0	6	4
oranges	6	0	2
bananas	0	4	16

(A) 0.7 (B) 0.8 (C) 16 (D) 20

Q5. What is printed?

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

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

Q6. 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><img src="example.svg"></body></html>'

@app.route("/example.svg")
def handler1():
    print("Y")
    return "TODO"

@app.route("/out.svg")
def handler2():
    print("Z")
    return "TODO"
```

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

Q7. If a BST is constructed using the algorithm we learned in class, and the insert order is `[6, 10, 5, 11]`, where will `11` be?

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

Q8. What is `x`?

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

    def __len__(self):
        return 5

    def __getitem__(self, lookup):
        return 1

obj = MyList([2, 4, 3])
x = len(obj.vals) # careful!
```

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

Q9. If you want to produce a dendrogram, which of the following should you use?

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

Q10. 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)`

Q11. The columns of a DataFrame contain weights in different units (pounds, kilograms, grams, etc). You want to group together similar rows. What would be a good transformer to use, prior to the clustering stage of the pipeline?

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

Q12. If you want to find the centroids of geographic regions, you should first convert to a coordinate reference system that is based on:

(A) degrees (B) latitude/longitude (C) meters (D) pixels

Q13. If `A=np.array([[6, 4], [1, 3]])` and `B=np.array([[5], [2]])`, what is `A*b`?

(A) `[[30, 20], [2, 6]]` (B) `[[30, 8], [5, 6]]` (C) `[[38], [11]]` (D) `[[42, 28], [7, 21]]`

Q14. If a single numeric feature is transformed by a `PolynomialFeatures(include_bias=True)`, how many output columns will there be?

(A) 1 (B) 2 (C) 3 (D) 4

Q15. Part of the computation performed by `LogisticRegression` for multi-class prediction involves finding the position of the largest number in each row of a matrix (let's call it `M`). How can this be done?

(A) `M.max(axis=0)` (B) `M.max(axis=1)` (C) `M.argmax(axis=0)` (D) `M.argmax(axis=1)`

Q16. Which of the following is NOT a `shapely` shape?

(A) Point, (B) Circle, (C) Polygon, (D) MultiPolygon

Q17. Given an array `X` of features, what calculation most closely corresponds to what `lr.predict` does, where `lr` was created with `LinearRegression(fit_intercept=False)`?

(A) `X * lr.coef_.reshape(-1,1)`
(B) `X @ lr.coef_.reshape(-1,1)`
(C) `X * lr.coef_.reshape(-1,1) > 0`
(D) `sigmoid(X @ lr.coef_.reshape(-1,1))`

Q18. How many oranges are classified as bananas, according to the following confusion matrix?

	apples	oranges	bananas
apples	83	85	1
oranges	27	26	61
bananas	13	43	54

(A) 18 (B) 43 (C) 61 (D) 104

Q19. Fill in the blank: every _____ has one instruction pointer associated with it.

(A) thread (B) process (C) program (D) computer

Q20. Given points `[(1, 5), (2, 4), (4, 3), (5, 2)]` and starting centroids `(2, 5)` for cluster 1 and `(4, 4)` for cluster 2, what points will be assigned to the cluster 1 after the first iteration of assigning points and updating centroids, using the k-means implementation from class?

(A) `[(1, 5), (2, 4)]` (B) `[(1, 5), (4, 3)]` (C) `[(2, 4), (4, 3)]` (D) `[(2, 4), (5, 2)]`

Q21. For KMeans, what is one generally hoping to have in the end?

(A) few clusters, low inertia
(B) few clusters, high inertia
(C) many clusters, low inertia
(D) many clusters, high inertia

Q22. The shape of `A` is (5, 3), the shape of `B` is (3, 9), and the shape of `C` is (9, 4). What is the shape of `A@B@C`?

(A) (3, 9) (B) (5, 3) (C) (5, 4) (D) (9, 4)

Q23. If you want to randomly split your data into train and test, but you don't want your results to change if you re-run your notebook, what should you pass to `train_test_split`?

(A) `test_size=0.75` (B) `train_size=320` (C) `random_state=50` (D) `stratify=False`

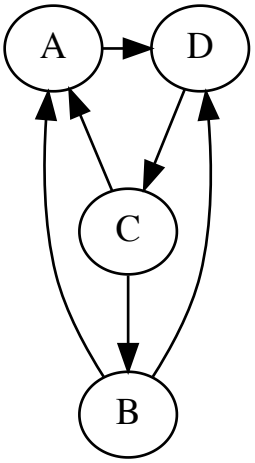
Q24. Generally speaking, what do you hope to get from `cross_val_score` for a regression?

(A) low `mean`, low `std` (B) low `mean`, high `std`, (C) high `mean`, low `std` (D) high `mean`, high `std`

Q25. Does the regular expression `r"b.*[A-Z]"` match anything in the string `"Bd.ba1c"`?

(A) yes (B) no

Q26. What can be said about the following graph?

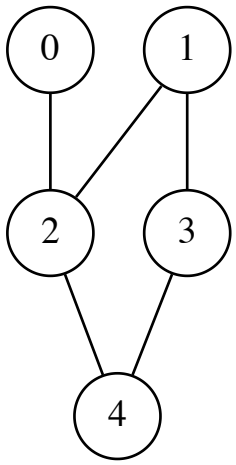


(A) it is not acyclic and not connected
(B) it is connected but not acyclic
(C) it is acyclic but not connected
(D) it is both connected and acyclic

Q27. The `explained_variance_ratio_` given by `p = PCA()` is `[0.8, 0.1, 0.07, 0.03]`. If you want to throw away some data while preserving 90% of the variance, what can you do to `p.components_`?

(A) `p.components_[:2, :]`
(B) `p.components_[2:, :]`
(C) `p.components_[:, :2]`
(D) `p.components_[:, 2:]`

Q28. What is the order in which the nodes of the undirected graph are visited in a BFS starting from node 0? When you have the choice of two or more nodes, break ties by choosing the node with smaller value.



(A) [0, 2, 0, 1, 4, 3] (B) [0, 2, 1, 3, 4] (C) [0, 2, 1, 4, 3] (D) [3, 0, 2, 4, 1]

Q29. `DataFrame` is in the `gpd.GeoDataFrame.__mro__` tuple. What does this imply?

(A) `GeoDataFrame` has all the methods `DataFrame` has
(B) `DataFrame` has all the methods `GeoDataFrame` has
(C) `GeoDataFrame` and `DataFrame` have an identical set of methods

Q30. What does `nums` contain after the following runs?

```
nums = []
def h(x):
    if x < 8:
        h(x+1)
    nums.append(x)
h(5)
print(nums)
```

(A) `[]` (B) `[8, 7, 6]` (C) `[6, 7, 8]` (D) `[8, 7, 6, 5]` (E) `[5, 6, 7, 8]`