## **Instructions** (shown before students start the test)

Remember to work on the Coding Challenges as well as the written test. You can work on both simultaneously. CAREFULLY NOTE THE INSTRUCTIONS FOR COMPLETING THE CODING CHALLENGES (it's the first problem).

You can use unlimited notes, but you obviously can't ask someone else for help. Good luck!!

## **Introduction** (shown after students start the test)

**1. (40.00 pts)**
CAREFULLY follow these instructions. Open the link below to the coding challenges. Click the button "Copy to Drive." This will create a copy of the Jupyter Notebook in your own Google Drive. THIS IS THE COPY YOU SHOULD EDIT. If you edit the original, your changes will not be saved.

Complete the coding challenges on your copy of the notebook. Add your team number to the name of the notebook. (You can click "UT Regional.ipynb" to edit the name.) If you're working with a partner, make sure you and your partner are working on the same copy (you will need to share your copy of the Notebook with your partner).

Before it's time to submit, go to the "Share" button (on your copy) and click "Change" under "Anyone on the internet with this link can view." Set the access privilege to "Editor" rather than "Viewer" for "Anyone with the link." Then click "Copy link" and submit it using the google form link below. Make sure you do this before time runs out!

Coding Challenges: https://colab.research.google.com/drive/11BP-pFd4LLznCKOpRAqVwx65NIxTGWAC?usp=sharing (https://colab.research.google.com/drive/11BP-pFd4LLznCKOpRAqVwx65NIxTGWAC?usp=sharing)

Google form link: https://forms.gle/am9twWsEky5vKtU89 (https://forms.gle/am9twWsEky5vKtU89)

**2. (2.00 pts)** Python uses an interpreter, which means

- A) The way the program is run depends on the computer’s interpretation
- B) Code isn’t compiled to a binary before being run
- C) The code has english-like syntax
- D) The programmer can easily understand the source code

**3. (2.00 pts)** Which of the following best describes inheritance?

- A) When one instance of an object chooses to reveal its internal field data to another.
- B) When a function has its default implementation set by another function.
- C) When a program uses lambda functions to pass data to smaller data structures.
- D) When one type of object acquires the properties/behaviors of another type.

**4. (2.00 pts)** A major difference between a list and a set is

- A) Sets are unordered, while lists are ordered.
- B) Lists are efficiently searchable, while sets are slow to search through.
- C) Lists don’t support duplicate entries, while sets do.
- D) Sets use key-value pairs while lists don’t.

**5. (2.00 pts)** A list comprehension is

- A) A technique for simplifying and reducing list data.
- B) Special "syntactic sugar" to create new lists.
- C) A system for documenting new list operations.
- D) A way to convert lists into other data structures.

**6. (2.00 pts)**
Sets are collections of many items, and yet it's a constant-time operation to find an element of a set (regardless of how big the set is). What technique makes this possible?

- A) Class polymorphism
- B) Public-key cryptography
- C) Hashtables
- D) Amortized analysis

**7. (2.00 pts)** Which of the following lines correctly instantiated an object in Python? Assume that the class MyObj has been defined already.

- A)
```
x = MyObj
```
- B)
```
x = MyObj()
```
- C)
```
x = MyObj.init()
```
- D)
```
x = new MyObj()
```

**8. (2.00 pts)** Which of the following code segments is free of both syntax errors and logic errors?

- A)
```
mylist = [(i, i+1) for i in range(10)]
count = 0
for ind in range(20):
    a, b = mylist[ind]
    count += a + b
print(count)
```
- B)
```
mylist = [(i, i+1) for i in range(10)]
count = 0
for a, b in mylist:
    count += a + b
print(count)
```
- C)
```
mylist = [(i, i+1) for i in range(10)]
count = 0
for a in mylist:
    count += a
print(count)
```
- D)
```
mylist = [(i, i+1) for _ in range(10)]
count = 0
for a, b from mylist:
    count += a + b
print(count)
```

**9. (2.00 pts)** What is the time complexity of finding the mean of a sorted list?

- A) Constant
- B) Logarithmic
- C) Linear
- D) Exponential

**10. (2.00 pts)** What is the time complexity of finding the median of a sorted list?

- A) Constant
- B) Logarithmic
- C) Linear

○ D) Exponential

**11. (2.00 pts)** What is the time complexity of binary search through an ordered list?

○ A) Constant

○ B) Logarithmic

○ C) Linear

○ D) Exponential

For the next 5 problems, consider the following code. Given a set of points in the Cartesian plane, it computes the perimeter of the polygon formed by the points. The points are given as two lists: one containing all the x-coordinates, and the other containing the corresponding y-coordinates. We'll assume that the points are in the correct order (so that connecting the points sequentially will yield a polygon).

```
def calc_perim(x_coords, y_coords):
    perim = 0
    N = len(x_coords)
    for i in range(N):
        dx = x_coords[i] - x_coords[i-1]
        dy = y_coords[i] - y_coords[i-1]
        perim += (dx**2 + dy**2)**(0.5)
    return perim
```

**12. (2.00 pts)** There's an error in the code. Which of the following would fix the error?

○ A) Changing

```
for i in range(N):
```

to

```
for i in range(N+1):
```

○ B) Changing

```
x_coords[i-1]
```

to

```
x_coords[(i-1)%N]
```

and likewise for

```
y_coords[i-1]
```

○ C) Change

```
perim += (dx**2 + dy**2)**(0.5)
```

to

```
perim += (dx^2 + dy^2)^(0.5)
```

○ D) Change

```
N = len(x_coords)
```

to

```
N = len(y_coords)
```

**13. (2.00 pts)** In principle, the two inputs x_coords and y_coords could be lists of different lengths. Which of these is the best practice for dealing with this possibility?

○ A) Change nothing; the algorithm will still work.

○ B) Return a pre-defined number (e.g. 25) if the list lengths don't match.

○ C) Print the difference in length between the two lists at the beginning of the function.

D) Use a try/except clause to raise an exception if the list lengths don't match.

**14. (2.00 pts)** Suppose, instead of two lists of coordinates, you wanted a single list of 2-tuples containing each (x, y) point. Which of the following lines would produce that list?

A)
```
coords = zip(x_coords, y_coords)
```

B)
```
coords = [(x, y) for x, y in x_coords, y_coords]
```

C)
```
coords = enumerate(x_coords, y_coords)
```

D)
```
coords = reduce(map(x_coords, y_coords), (x, y))
```

**15. (2.00 pts)** What is the time complexity of the algorithm?

A) Constant
B) Logarithmic
C) Linear
D) Quadratic

**16. (2.00 pts)** What is the space complexity of the algorithm? Don't include the space used by the input; only consider the space allocated by the algorithm itself.

A) Constant
B) Logarithmic
C) Linear
D) Quadratic

**17. (2.00 pts)** Which of the following is NOT true about the variance of a probability distribution?

A) It roughly correlates with the width of the probability distribution function
B) It measures how far the mean is from 0
C) It measures deviations from the mean
D) It cannot be zero for a smooth, continuous probability distribution.

**18. (2.00 pts)**
Suppose $p_1(x)$ is a Gaussian probability distribution function with mean $\mu_1$ and $p_2(x)$ is a Gaussian probability distribution function with mean $\mu_2$. Both distributions have the same variance. Consider a new distribution function $p_3(x) = \frac{1}{2}(p_1(x) + p_2(x))$. What is the mean of $p_3(x)$?

A) Need more information.
B) $1/2$
C) $\mu_1 + \mu_2$
D) $\frac{1}{2}(\mu_1 + \mu_2)$

**19. (2.00 pts)** Covariance is a measure of

A) How distant two quantities are from each other in feature space
B) How closely two quantities are correlated
C) The variance of a quantity with respect to itself
D) How skewed a multivariate distribution is, relative to its mean

**20. (2.00 pts)** The parameters required to fully specify a normal distribution are:

A) Mean only

B) Mean and variance only

C) Variance and kurtosis only

D) Mean, variance, and kurtosis

**21. (2.00 pts)** What is the difference (in statistical terms) between your GPA and your SAT superscore?

(Your SAT superscore is the sum of your highest score on each test section. It could combine section scores from different times you took the SAT.)

A) One has a well-defined variance while the other doesn't.

B) One is a probability distribution while the other isn't.

C) One has a well-defined median while the other doesn't.

D) One is a measure of central tendency while the other isn't.

**22. (7.00 pts)**

In order to solve a classification problem, Reya is testing out two learning algorithms: RF and Naïve Bayes. The testing bank consists of 100 images to classify, and each algorithm is run a total of 8 times on the dataset. The number of correctly classified images for each algorithm in each trial is recorded below:

RF: 91, 92, 90, 83, 85, 88, 94, 83

Naïve Bayes: 78, 78, 81, 90, 76, 83, 80, 84

Using this information, Reya believes that RF is better than Naïve Bayes for this classification task. In order to mathematically prove this, she needs your help in conducting a two-sample t-test. Answer the questions below.

1. Express the null hypothesis and alternate hypothesis.

2. Calculate the t-statistic for this test. Show your work (write the formula).

3. Using a significance value of alpha = 0.05, calculate the p-value. What can you conclude about this hypothesis test (and the experiment)?

**23. (7.00 pts)**

At your job at the zoo, you are given the task of classifying different images of your residents. Your boss instructs you to create some sort of machine learning model. Answer the questions below regarding your task.

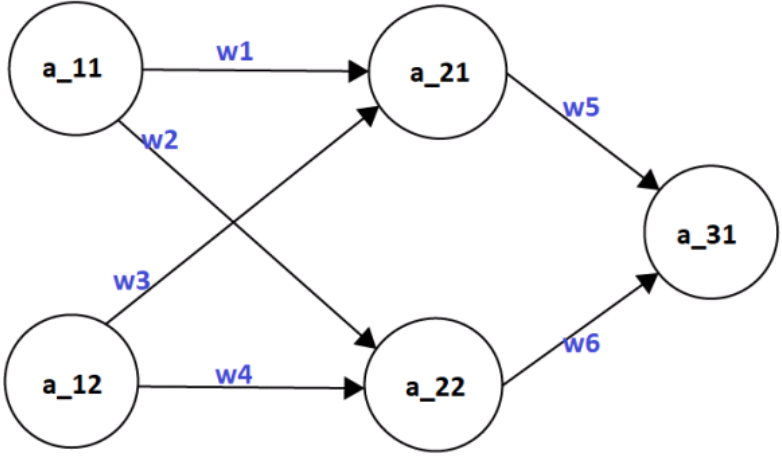
1. Would this be an unsupervised or supervised learning question? Briefly explain.

2. What type of neural network is best fit for this task? Briefly explain the concept behind this type of neural network.

3. Using your dataset of animal pictures, you learn about batch, mini-batch, and stochastic gradient descent. What are the key differences between these methods?

4. At your internship at the weather station, your manager has placed you on the difficult task of predicting the weather. You plan to use a sort of random algorithm, which places emphasis on previous states (the weather a few days before) in order to make a prediction. What's the name of this approach?

**24. (8.00 pts)**




For the next question, refer to the diagram above.

You and Sanjay are having a discussion about deep learning at lunch. Sanjay, being the smart aleck he is, states that only computers can calculate values in a neural network; however, you want to prove him wrong using the simple example above. For simplicity, assume that there are no biases and that this network uses the linear activation function A(x) = cx + d (where c and d are constants). Then, answer the questions below.

1. Express the activated value $a_{21}$ in terms of the two inputs, $a_{11}$ and $a_{12}$. You do not need to expand your answer.

2. Given $a_{21} = 0.1$, $a_{22} = 0.3$, $w_5 = 2$, $w_6 = 5$, c = 3, and d = 5, what is the value of the activated output ($a_{31}$)?

3. In real-world neural networks, there are many more layers/nodes than depicted in the photo. Since computing values for each node individually would be inefficient, what technique do computers employ to optimize operations?

We hope you enjoyed this exam! If you have any feedback about any of the exams at this tournament, please let us know through this form: https://tinyurl.com/utreg21feedback (https://tinyurl.com/utreg21feedback)