**CIST 360/557**

# Homework 2: NumPy Practice

### Thinking Like a Machine Learning Model

## Instructions

This assignment helps you understand how data is structured for machine learning models using NumPy. Each question builds your intuition for array shapes, operations, and data organization.

- Answer each question concisely (1-2 sentences or an expression).
- For code questions, write the NumPy expression or indexing notation.
- Each sub-question is worth 1 point.
- Total: 20 points

**Q1. Smartwatch Data Collection (3 points)**

A smartwatch collects the following data every day:

- steps walked
- hours of sleep
- average heart rate

a) How many features are recorded per day?

**ans:// We are recording 3 features (steps, hours of sleep and average heart rate).**

b) If the watch collects data for 365 days, what should be the shape of the NumPy array used to store this data? (Assume each day is one row and each feature is one column.)

**ans:// The NumPy array will be (365, 3), where 365 are the samples and 3 are the features. (n_sample, n_features)**

c) What does one row represent?

**ans:// One row represents day (a sample of a day, it could be any way).**

**Q2. Grayscale Photo Storage (3 points)**

A grayscale photo is stored as a NumPy array with shape (28, 28).

a) How many total values are stored in this array?

**ans:// Values are (n_samples x n_features), so 28 * 28 = 784**

b) If you have a single photo with shape (28, 28), what shape transformation is needed before it can be used as input for a machine learning model that expects 2D input (samples × features)?

**ans:// if we have 1 single photo and has a shape of 28 x 28 we will have a (1, 28*28) = (1, 784) 2D input**

c) What would the shape be if you have 10,000 such photos stored in one array?

**ans:// If the shape is (1, 784) and we have 10,000 photos (samples) we get → (10000, 784) array.**

### Q3. Fitness App Calibration (1 point)

A fitness app wants to increase all users' daily step counts by 10% due to a sensor calibration update. Given an array

`steps` containing daily step counts, write this operation as a single NumPy expression (no loops needed).

**ans:// The NumPy expression is:**

```python
steps = steps * 1.10
```

### Q4. Exam Score Normalization (3 points)

You have a NumPy array of exam scores with shape (200, 3):

- column 0: quiz score
- column 1: midterm score
- column 2: final exam score

You subtract the average score of each column from the entire array.

a) When you subtract a 1D array of shape (3,) containing the column means from a 2D array of shape (200, 3), what does NumPy broadcasting do? Explain how the subtraction is applied.

**ans:// NumPy broadcasting will reshape the 1D array of (3,) → (1,3),**

**Then NumPy will broadcast this row down the 200 rows. Finally each row subtracts the same column-wise means.**

b) After this subtraction, what does a value of 0 mean for a particular student's score?

**ans:// A value of 0 for a particular student's score means that the student score is equal to the average score.**

c) Why might this operation (called "mean centering") be useful before training a machine learning model?

**ans:// Mean centering might be useful because it puts each feature around 0 and so the model can learn more efficiently.**

### Q5. Axis Confusion (3 points)

A food delivery app records delivery times (in minutes) for 500 orders over 7 days. The data is stored as an array with shape (500, 7).

a) What does axis=0 represent in this array?

**ans:// axis=0 represents the rows in the (500, 7) array. So the 500 orders.**

b) What does axis=1 represent in this array?

**ans:// axis=1 represents the columns in the (500, 7) array. So the 7 days.**

c) If you compute `np.mean(delivery_times, axis=1)`, what question are you answering about the delivery data? What does each value in the result represent?

**ans:// The question being answered about the delivery data is: what is the average time of each order across the 7 days?**

**And each value in the results represents the mean delivery time for a specific order.**

### Q6. Data Type Issues (1 point)

You accidentally store numbers and strings together in a NumPy array. For example, if you create `arr = np.array([1, 2, 'three', 4])`, what happens to the dtype of the array?

**ans:// the dtype of the array will become a string (Unicode) type.**

```python
array(['1', '2', 'three', '4'], dtype='<U21')
```

### Q7. Reshaping Without Changing Meaning (2 points)

A streaming platform records how many minutes each user watches per day. You have a NumPy array of shape (300,).

a) What does this array represent? (How many samples? How many features per sample?)

**ans:// the array represents the number of users in the streaming platform. There is only 1 feature per sample that represents the minutes watched per day.**

b) How would you reshape it to shape (300, 1)? Write the code.

```python
WatchTime = WatchTime.reshape(300, 1)
```

**Q8. Data Selection (Indexing) (4 points)**

A music streaming app stores listening data in a NumPy array with shape (1000, 5):

- Each row represents one user
- Each column represents a feature:

column 0: total minutes listened

column 1: number of songs played

column 2: number of playlists

column 3: skips

column 4: likes

***Note:*** *Python uses 0-based indexing, so the first user is at index 0.*

a) How do you select all the data for the 10th user? Write the indexing code.

Answer:

```python
""" lets say our NumPy array is called MusicData, To select the 10th """

MusicData[9]
```

b) How do you select the "skips" feature (column 3) for all users? Write the indexing code.

```python
""" Where the : means select all users and 3 means select the skips column """

MusicData[:, 3]
```

c) How do you select only the "songs played" (column 1) and "likes" (column 4) features for all users, returning them together in one array? Write the indexing code.

```python
""" The nested [1, 4] means select columns 1 (songs played) and 4 (likes) and
the outer [:] means select all users"""

MusicData[:, [1, 4]]
```

d) What is the shape of each result from parts a), b), and c)?

**ans:// The shape of a) is (1,1)**

**The shape of b) is (1000,1)**

**The shape of c) is (1000, 2)**

---

***End of Homework 2***

*Total Points: 20*