Classroom 1: Why are .csv files commonly used to import data?

- **A) Because they store data in a structured table format that's easy for programs to read**
- B) Because they are simple to upload and download, making them convenient for many projects
- C) Because they are encrypted to protect the data from hackers
- D) Because .csv files are the only format available for importing datasets

Hint 1: A .csv file organizes data using rows and columns, which makes it easy for code to scan through it efficiently.

Hint 2: Even though it's just plain text, .csv files use a structured format that lets tools like Excel or Python easily understand the data

Explanation: A .csv file stores information by using the first line to store the column headers, with each column separated by commas. Each subsequent line includes a datapoint that corresponds to its respective column.

Classroom 2: Why is it important to gather as much of the data as possible when working with a dataset?

- A) Because some types of data are harder to predict, increasing the time needed for analysis
- **B) Because incomplete data can lead to inaccurate predictions and misleading results**
- C) Because datasets are only valid if every single value is known
- D) Because missing values are automatically discarded, which corrupts the entire dataset

Hint 1: Without seeing the full picture, it's much harder to make the right decision—just like trying to solve a puzzle with missing pieces.

Hint 2: Even if most of the data is present, missing just a few important values can still throw off your results.

Predicting or filling in missing data can introduce errors, especially if the missing values belong to complex or hard-to-predict categories. For example, if the original diagnosis data includes two benign and one malignant case, but a prediction mistakenly swaps them (e.g., two malignant and one benign), the outcome becomes misleading. That's why having as much accurate data as possible is crucial for effective analysis and reliable conclusions.

Classroom 3: What is a univariate model?

A) A model that uses two or more variables to make predictions
B) A model that analyzes only one variable at a time
C) A model that ignores data and makes random guesses
D) A model used only for images and sound data

Hint 1: Think of the word "uni" — it usually means "one", like in "unicycle" or "universe".

Hint 2: Univariate models are the simplest kind — they don't look at relationships between multiple columns, just focus on a single one.

Explanation: A univariate model focuses on a single variable — for example, analyzing just the "concave points_mean" column from a dataset without using any other data. This type of model helps us understand or predict outcomes based on only one feature, making it useful for simple patterns or when data is limited. It's a good starting point before moving on to more complex models that use multiple variables.

Classroom 4: What do the dark red bars represent in these histograms?

A) Areas where neither benign nor malignant cases are present
B) Data points that are equally spaced between histogram bins
C) Overlapping values where both benign (B) and malignant (M) diagnoses occur
D) Errors in data collection that result in color blending

Hint 1: Think about what happens when two different colored transparent bars stack on top of each other in a histogram.

Hint 2: The graph compares two groups (B and M). If they share the same value range, what would that look like visually?

Explanation:The dark red area represents the overlap between the two distributions. Where the bars for each class (B and M) overlap at the same bin range indicating that both B and M samples fall within that specific value range.

Classroom 5: In a heatmap showing correlation values, what does it mean when a number is greater than 0 and approaches 1, and what does it mean when it's less than 0 and approaches -1?

A) A number close to 1 means strong positive correlation, while a number close to -1 means strong negative correlation
B) A number close to 1 means the data is corrupted, and a number close to -1 means it's missing
C) It means both variables are becoming identical as the number gets closer to 1 or -1

D) Numbers close to 1 or -1 always mean the values are wrong and should be removed

Hint 1: A positive number close to 1 means that as one value increases, the other tends to increase too.

Hint 2: A negative number close to -1 means that when one value increases, the other tends to decrease — like they move in opposite directions.

Explanation:

In a heatmap of correlation values, numbers range from -1 to 1.

A value close to 1 indicates a strong positive correlation — both variables increase together.

A value close to -1 shows a strong negative correlation — when one goes up, the other goes down.

A value around 0 means there's no clear relationship between the two variables.

Heatmaps make it easier to visually spot relationships between features in a datase