Confusion Matrix

1. Focus on the names:
    a. True: Correct prediction (the model's guess was right)
    b. False: Incorrect prediction (the model's guess was wrong)
    c. Positive: Positive class (for example, "5")
    d. Negative: Negative class (for example, "not 5")
2. Create a simple scenario
    a. True Positive: You correctly identified as 5 as a 5.
    b. True Positive: You correctly identified a non-5 as not 5.
    c. False Positive: You incorrectly identified a non-5 as a 5. (a false alarm)
    d. False Negative: You incorrectly identified a 5 as not 5. (you missed it).
3. Use a story:

    Create a little story to remember:

    - True Positive: "I saw a 5 and correctly called as a 5."
    - False Positive: "I mistook something that wasn't a 5 for a 5." (You saw something that wasn't there, a positive mistake)
    - True Negative: "I saw something that wasn't a 5 and correctly said it wasn't")
    - False Negative: "I missed a 5 and said it wasn't a 5." (You missed something that was there, a negative mistake.)
4. Visualize:

    Create a mini table and repeatedly write down the meanings of each cell. For example:

	Predicted 0	Predicted 1
Actual 0	True Negative	False Positive
Actual 1	False Negative	True Positive

1. One-Versus-the-Rest (OvR):
    a. How does it work?
        i. You train a separate model for each digit. For example, "Is it 0 or something else?"
    b. What does it do?
        i. You create 10 different models (each trying to recognize one digit)
    c. How is the result found?
        i. To classify an image, you get results from all models and choose the digit with the highest score.
    d. Example:
        i. A 0 detector, a 1 detector, a 2 detector, ... and so on.
2. One-Versus-One (OvO):
    a. How does it work?
        i. You train a model for every pair of digits. For example, "Is it 0 or 1?", "0 or 2?", etc.
    b. What does it do?
        i. If there are 10 classes (0-9), you train 45 different models (one for each digit pair).
    c. How is the result found?
        i. To classify an image, you run it through all 45 models, and the digit that wins the most 'duels' is chosen as the result.
    d. Example:
        i. A model to separate 0 and 1, another model for 1 and 2 ... and so on.

Main differences

- OvR: You train fewer models (10 total), but each model uses the entire dataset.
- OvO: You train many models (45 total), but each model only needs to distinguish two digits.

When is each strategy preferred?

- If your algorithm slows down with large datasets, you might prefer OvO because it works with smaller subsets of the data.
- Usually, OvR is preferred because it requires fewer models to be trained.

In summary, both strategies provide different ways of classifying digits. OvR separates each class one at a time, while OvO compares each pair of classes.