Context:

In a study, we analyze two different medical conditions based on systolic blood pressure levels, which are assumed to follow normal distributions.

- **Condition A (e.g., Hypertension):**
  - Mean blood pressure: $\bar{x}_1 = 140$ mmHg
  - Standard deviation: $\delta_1 = 15$ mmHg
  - Probability density function:

$$y_1 = \frac{1}{\sqrt{2\pi\delta_1^2}} e^{-\frac{(x-\bar{x}_1)^2}{2\delta_1^2}}$$

- **Condition B (e.g., Normotension):**
  - Mean blood pressure: $\bar{x}_2 = 120$ mmHg
  - Standard deviation: $\delta_2 = 10$ mmHg
  - Probability density function:

$$y_2 = \frac{1}{\sqrt{2\pi\delta_2^2}} e^{-\frac{(x-\bar{x}_2)^2}{2\delta_2^2}}$$

- **Part C: Random Sample Generation and Error Calculation**
- **Question:** Using Python, generate two random samples of 500 patients for each condition based on the defined normal distributions. Calculate the classification errors based on the decision threshold obtained in Part B and compare these errors with the analytical estimates calculated in Part B.

Part D: Second Feature - Cholesterol Levels

Question: Consider a second feature: total cholesterol levels for each patient, which also follows a normal distribution.

- **Cholesterol for Condition A:** Mean: $\bar{C}_1 = 240$ mg/dL, Standard deviation: $\delta_1 = 30$ mg/dL
- **Cholesterol for Condition B:** Mean: $\bar{C}_2 = 180$ mg/dL, Standard deviation: $\delta_2 = 20$ mg/dL

Repeat Parts A, B, and C for this new feature (cholesterol levels).

Part E: Scatter Plot of Features

Question: Calculate both features (blood pressure and cholesterol levels) for 500 patients from each condition and create a scatter plot to visualize the data distribution.

matlab Copy code

```
% MATLAB code snippet for scatter plot
features = [hypertension_data; normotension_data];
cholesterol_levels = [cholesterol_conditionA; cholesterol_conditionB];
labels = [zeros(500, 1); ones(500, 1)]; % 0: Condition A, 1: Condition B
```

Part F: Classification Using a Perceptron

Question: Using a perceptron, perform classification under different training/testing scenarios:

- 250 for training and 250 for testing
- 300 for training and 200 for testing
- 350 for training and 150 for testing
- 400 for training and 100 for testing
- 450 for training and 50 for testing

Discuss whether the error rates can be computed analytically and how they compare to the empirical results from the perceptron.