Machine Learning Basics - Set 2:
Decision Trees, Fisher Discriminant and Feature Transformation

Datasets: A (from Set 1), B: Similar to dataset A, but not linearly separable unless we apply a transformation before.

1. **Two-layer Decision Tree:**
   a) Visualize Dataset A and propose two sequential cuts to separate the classes.
   b) Implement your two-layer decision tree and classify each data point.
   c) Calculate and report the signal efficiency and background rejection efficiency.
   d) Discuss how you decided on the order of your cuts and potential limitations of this approach.

2. **Fisher Discriminant:**
   a) Calculate the mean vectors $\mathbb{E}(\boldsymbol{x}|\boldsymbol{S})$ and $\mathbb{E}(\boldsymbol{x}|\boldsymbol{B})$ for signal and background classes in Dataset A.
   b) Compute the sum of variances matrix $\mathbb{V} = \mathbb{V}(\boldsymbol{x}|\boldsymbol{S}) + \mathbb{V}(\boldsymbol{x}|\boldsymbol{B})$.
   c) Calculate the Fisher Discriminant vector $\mathbf{w} = (\mathbb{E}(\boldsymbol{x}|\boldsymbol{S}) - \mathbb{E}(\boldsymbol{x}|\boldsymbol{B}))^2/\mathbb{V}$, which minimises $J(\mathbf{w})$.
   d) Project the data points onto the Fisher Discriminant axis and visualize the results with an histogram.

3. **Fisher Discriminant Classification:**
   a) Determine a suitable threshold for classification based on the projected data.
   b) Implement the Fisher Discriminant classifier and evaluate its performance on Dataset A.
   c) Compare the performance of the Fisher Discriminant to the methods used in Set 1.

4. **Feature Transformation:**
   a) Analyze Dataset B and identify why the standard Fisher Discriminant might fail.
   b) Propose a feature transformation that could improve class separation. Justify your choice.
   c) Implement your feature transformation and visualize the transformed data.
   d) Apply the Fisher Discriminant to the transformed data and compare its performance to the non-transformed case.

5. ***Advanced Analysis:**
   a) Implement a function to perform k-fold cross-validation (with k=5) for your classifiers.
   b) Use cross-validation to compare the robustness of the decision tree, original Fisher Discriminant, and transformed Fisher Discriminant methods.
   c) Discuss the strengths and weaknesses of each method based on your cross-validation results.
   d) Implement Principal Component Analysis (PCA) as an automated feature transformation method:
      - Calculate the covariance matrix of the dataset.
      - Compute the eigenvectors and eigenvalues of the covariance matrix.
      - Sort the eigenvectors by decreasing eigenvalues and select the top k eigenvectors.
      - Project the data onto the selected eigenvectors.
   e) Apply PCA to Dataset B, reducing it to 2 principal components. Visualize the transformed data.
   f) Use the Fisher Discriminant on the PCA-transformed data. Compare its performance to your previous results.
   g) Discuss the benefits and limitations of using PCA for feature transformation in this context.

Note: If you complete these problems quickly, feel free to move on to Set 3.