Machine Learning Basics - Set 1:
Simple and Linear Classification

Dataset A: A linearly separable dataset with two features and points belonging either to class 0 or 1.

1. **Data Visualization:**
   a) Plot the given 2D dataset, distinguishing between signal and background classes.
   b) Describe any patterns or separations you observe in the data.
2. **Simple Cut-based Classification:**
   a) Propose a classification rule of the form $y < y_{cut}$. Justify your choice of y_{cut}.
   b) Implement your rule and classify each data point.
   c) Calculate and report the signal efficiency and background rejection efficiency.
3. **Linear Classification:**
   a) Propose a linear classification boundary $y = kx + l$. Explain how you determined k and l.
   b) Implement your linear classifier and classify each data point.
   c) Calculate and report the signal efficiency and background rejection efficiency.
   d) Compare the performance of this linear classifier to the simple cut-based method. Discuss advantages and limitations of each approach.
4. **Performance Evaluation:**
   a) Implement a function to calculate the Area Under the Curve (AUC) for your classifiers.
   b) Create Receiver Operating Characteristic (ROC) curves for both the cut-based and linear classifiers.
   c) Compare the AUC values. Which classifier performs better and why?
5. ***Optimization Challenge:**
   a) Perform a 2D grid search to find the optimal values of (k, l) for the linear classifier based on AUC.
   b) Implement your optimized linear classifier and evaluate its performance.
   c) Discuss the trade-offs between computational cost and performance improvement in this optimization process.

Note: If you complete these problems quickly, feel free to move on to Set 2.