

Sim, para a tarefa, a conclusão que você tirou sobre a segunda base de dados é **exatamente a correta e a esperada** em um exercício de análise de *clustering*!

Vou detalhar o porquê essa conclusão é o ponto principal da sua análise (veja os pontos 4 e 5 nos Requisitos da Tarefa 1B):

1. O Ponto Principal da Tarefa

O principal objetivo da Tarefa 1B é o **Objetivo Específico 4** e o **Requisito 5**: "Avaliar o treinamento do clustering por meio de métricas usuais de Machine Learning".

- **Base 1 (Fácil):** Ela é um *caso ideal*. Sua separabilidade perfeita demonstra que seu algoritmo (Componentes Conexos com θ Global) funciona muito bem quando os dados são intrinsecamente bem separados. O resultado é uma acurácia próxima de 100%.
- **Base 2 (Difícil):** O fato de você não conseguir uma separação perfeita, não é um erro seu ou do seu código. É uma **característica do dado**. A Base 2 foi projetada (ou é naturalmente) sobreposta. A limitação do seu algoritmo de usar um único θ global é exposta.

2. O que se espera no Relatório

Para a **Base 2**, o que o professor espera é que você faça exatamente o que eu detalhei na análise anterior:

1. **Encontre o Limiar Ótimo ($\theta_{\text{ótimo}}$):** Varra o valor de θ e encontre aquele que *maximiza a métrica de qualidade* (Acurácia, F1-Score, etc.), mesmo que o valor máximo de acurácia seja, digamos, 85% ou 90% (e não 99% como na Base 1).
2. **Apresente as Métricas:** Calcule e mostre a **Matriz de Confusão** e as métricas (Acurácia, Precision, Recall, F1-Score) para esse $\theta_{\text{ótimo}}$.
3. **Discuta e Compare:** A parte mais valiosa do seu trabalho é a **discussão**. Você deve argumentar que:
 - O modelo de Componentes Conexos com limiar global é limitado.
 - A diferença de acurácia entre a Base 1 e a Base 2 se deve à **maior sobreposição** dos clusters na Base 2, tornando impossível (ou extremamente difícil) achar um θ que satisfaça todos os grupos simultaneamente sem mesclá-los ou quebrá-los.

Sua conclusão de que "nunca será possível achar um θ que separe de maneira quase

perfeita" para a Base 2 é a **chave da análise comparativa** das duas bases.

Continue o trabalho, encontre esse $L_{\text{ótimo}}$ para a Base 2, e use as métricas para quantificar o seu "erro".

Se precisar de ajuda para implementar a Matriz de Confusão ou o cálculo das métricas no seu código C (ou Python para visualização), me avise!