1. Introduction

Kính chào thầy cô và các bạn, hôm nay nhóm chúng em xin trình bày về **Principal Component Analysis (PCA)** – một phương pháp quan trọng trong xử lý dữ liệu và học máy.

📌 **Vấn đề đặt ra:** Trong thực tế, dữ liệu thường có **quá nhiều biến số (dimensions)**, làm cho việc phân tích trở nên khó khăn, mô hình học máy dễ bị quá khớp (overfitting) và tốn nhiều tài nguyên tính toán.

📌 **Giải pháp:** PCA là một kỹ thuật giúp **giảm số chiều dữ liệu**, nhưng vẫn giữ lại **lượng thông tin quan trọng nhất**, giúp dữ liệu dễ hiểu hơn và tăng hiệu quả tính toán.

📌 **Ứng dụng thực tế:**

✅ **Nhận diện khuôn mặt** trong trí tuệ nhân tạo

✅ **Giảm nhiễu** trong dữ liệu hình ảnh

✅ **Tăng tốc độ học máy** bằng cách giảm số lượng biến số

Trong bài thuyết trình này, chúng em sẽ giải thích **cách PCA hoạt động**, các **bước thực hiện**, và một số **ứng dụng thực tế** để mọi người hiểu rõ hơn về kỹ thuật này. Hy vọng bài trình bày sẽ mang lại nhiều thông tin hữu ích! 🚀

1. Purpose

Mục tiêu chính của PCA là **giảm số chiều của dữ liệu** trong khi vẫn giữ lại **phần lớn thông tin quan trọng**. Điều này giúp:

✅ **Đơn giản hóa dữ liệu**: Biểu diễn dữ liệu có nhiều biến dưới dạng một tập hợp nhỏ hơn nhưng vẫn giữ được xu hướng chính.

✅ **Giảm nhiễu và loại bỏ dư thừa**: PCA giúp loại bỏ các biến có mối tương quan cao, làm cho dữ liệu rõ ràng hơn.

✅ **Tăng hiệu suất mô hình học máy**: Giảm số lượng biến giúp mô hình học nhanh hơn, giảm nguy cơ quá khớp (overfitting).

✅ **Dễ dàng trực quan hóa dữ liệu**: Khi dữ liệu nhiều chiều được giảm xuống 2 hoặc 3 chiều, chúng ta có thể dễ dàng vẽ biểu đồ và phát hiện xu hướng.

Với những lợi ích này, PCA là một công cụ quan trọng trong phân tích dữ liệu, đặc biệt trong các lĩnh vực như **thị giác máy tính, xử lý ảnh, tài chính và sinh học**. 🚀

1. Problem Statement
2. Method
3. Geometrical Explanation
4. Model Checking
5. Case Study