Chào thầy và các bạn, hôm nay em sẽ làm video nói về thuật toán backprogation hay được gọi là thuật toán lan truyền ngược trong neural network, vì clip chỉ giới hạn từ 2 đến 5 phút nên em sẽ tập trung nói về các khái niệm, các bước chính trong thuật toán và ví dụ đơn giản và coi như thầy và các bạn đã biết và hiểu được mạng noron như thế nào, các lớp layer có trong mạng và cách tính độ lỗi. Trước hết nói về nhiệm vụ, mục đích của thuật toán lan truyền ngược là tìm cách để giảm độ lỗi của mô hình sao cho đầu ra của mô hình gần với giá trị mong muốn nhất sau từng bước học . Vậy ta sẽ giảm độ lỗi bằng cách nào thì thứ nhất, độ lỗi là một hàm số của các tham số weight và bias và nhiệm vụ để giảm cái loss này là tìm các tham số weight và bias mới tức là điều chỉnh weight và bias mới sao cho cái hàm loss này sẽ giảm xuống. Giả sử giá trị hàm số nằm ở đây và ta muốn giảm giá trị này tức là điều chỉnh xuống tức là đi theo cái đường cong biểu diễn hàm số đấy, để giảm giá trị của một hàm số xuống đến cái cực trị ở đây thì chúng ta phải hướng các tham số biểu diễn đường cong này từ từ hướng về cực trị để tối ưu hóa sao cho tối thiểu độ lỗi, quá trình này được gọi là giảm độ dốc ( Gradient Descent ). Vậy làm sao để giảm độ dốc dần dần về cực trị tức là ta phải tìm ra độ dốc mới và phương pháp chính mà ta nghĩ ngay đến chính là sử dụng đạo hàm, một phương pháp cơ bản để tìm được độ dốc giữa đường tiếp tuyến và đường cong tại một điểm cụ thể, tức là ta sẽ tính đạo hàm từng phần của hàm loss function trên các weight và bias. Sau khi tính được đạo hàm ta sẽ điều chỉnh lại weight và bias theo công thức sau w = w-axdentaw, và ta cần lưu ý rằng việc điều chỉnh weight và bias phụ thuộc vào độ lớn của alpha này alpha nhỏ thì bước điều chỉnh nhỏ thì độ lỗi của hàm sẽ thay đổi chậm và ngược lại, ta nên chọn alpha nhỏ vì nếu chọn giá trị alpha quá lớn có thể khiến giá trị mới vược qua cực trị vì vậy ta cần bước nhảy nhỏ để giá trị đạt tối thiểu cực trị chứ và giá trị alpha này được gọi là tốc độ học