

Chương 2

LÝ THUYẾT NỀN TẢNG

2.1 Lý thuyết nền tảng của mô hình tạo sinh đối nghịch

2.1.1 Kiến thức về mạng nơron máy học

- Mạng nơron đa lớp được tạo thành bởi nhiều nơron được tổ chức thành các lớp. Mạng nơron đa lớp thường có 3 thành phần: lớp đầu vào (Input layer), lớp ẩn (Hidden layer), lớp đầu ra (Output layer). Mạng nơron đa lớp thường được gọi là mạng nơron truyền thẳng (feedforward neural network). Mục tiêu của mạng nơron đa lớp là ước lượng các hàm f^* . Ví dụ, đối với bài toán phân lớp, $y = f^*(x)$ gán đầu vào x đến nhóm y . Một mạng nơron đa lớp sẽ định nghĩa một hàm $y = f(x; \theta)$ và học giá trị của tham số θ để được kết quả là một hàm tương đương tốt nhất. Mạng nơron đa lớp đóng vai trò quan trọng trong máy học, nó có mặt trong nhiều ứng dụng thương mại. Điển hình như mạng tích chập (convolutional networks) dùng để nhận dạng vật thể trong hình ảnh là một dạng đặc biệt của mạng nơron đa lớp.

- Lan truyền ngược: Quá trình sử dụng mạng nơon truyền thẳng để nhận một đầu vào x và sinh ra một đầu ra \hat{y} được gọi là lan truyền thẳng (forward propagation). Trong quá trình huấn luyện, lan truyền thẳng có thể tiếp tục cho tới khi nó tạo ra một chi phí $J(\theta)$. Thuật toán lan truyền ngược (back-propagation) cho phép thông tin từ chi phí chảy ngược lại trong mạng, để tính toán độ dốc (gradient).
- Không gian ẩn (Latent space) là một không gian mà các điểm dữ liệu tương đồng nhau thì sẽ ở gần nhau hơn.
- Mã ẩn (Latent code): là một vectơ đầu vào chứa các dữ liệu nén.

2.1.2 Mạng tạo sinh đối nghịch GAN

- Generator
- Discriminator

2.1.3 Mạng nơon tích chập CNN

- Filter layer
- Pool layer (Avg Pool, Global Avg Pool, Max Pool)
- Residual Block (Skip connection)
- Fully connected layer

2.1.4 Hàm kích hoạt Activation function

- ReLU
- Leaky ReLU
- Softmax

2.1.5 Thuật toán tối ưu

- Gradient descent
- Stochastic gradient descent
- Adam

2.1.6 Xác suất thống kê trong bài toán máy học

- Phân phối tự nhiên (Gaussian distribution) : Phân phối này có ý nghĩa quan trọng và được áp dụng rộng rãi trong cuộc sống. Ví dụ: nếu chúng ta lấy ngẫu nhiên chiều cao của 100 người, sắp xếp từ giá trị thấp nhất đến cao nhất và vẽ thành biểu đồ với trục hoành là chiều cao, trục tung là số lượng người có chiều cao đó thì sẽ được một biểu đồ dạng hình chuông. Đây cũng chính là dạng biểu đồ của phân phối chuẩn, biểu đồ hình chuông. Về công thức, cho một biến x thuộc vào phân phối chuẩn với giá trị trung bình và độ lệch chuẩn, chúng ta sẽ có hàm mật độ phân phối xác suất sau đây:

$$P(X = x | \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right]$$

- Phân phối đồng nhất (Uniform distribution): Phân phối đồng nhất là một dạng phân phối trong đó tất cả kết quả có khả năng xảy ra là như nhau. Ví dụ: trong một bộ bài, tỉ lệ rút được một lá bài có hình trái tim hoặc rút được lá bài có hình kim cương là như nhau. Việc tung một đồng xu hai mặt cũng có phân phối đồng nhất vì tỉ lệ mặt trên là mặt ngửa hay mặt sấp là như nhau. Đồ thị biểu diễn của phân phối đồng nhất sẽ là một đường thẳng nằm ngang.
- Entropy: Trong lý thuyết thông tin, entropy của một biến ngẫu nhiên là mức trung bình của "thông tin", "sự bất ngờ", hoặc "sự không

chắc chắn" vốn có trong các kết quả có thể có của biến đó. Khái niệm entropy thông tin được Claude Shannon đưa ra trong bài báo "Lý thuyết toán học về truyền thông" năm 1948. Ví dụ: một đồng xu không đồng nhất (biased coin) được tung lên có xác suất mặt ngửa nằm trên là p , mặt sấp nằm trên là $1-p$. Thì lúc này sự không chắc chắn lớn nhất khi $p = \frac{1}{2}$ khi không có chứng cứ nào xác định mặt trên là mặt ngửa hay mặt sấp, entropy lúc này là bit 1. Sự không chắc chắn nhỏ nhất khi $p = 0$ hoặc $p = 1$ khi chúng ta đã biết chắc mặt nào của đồng xu sẽ nằm ở trên, entropy lúc này là bit 0. Khi p thay đổi sẽ dẫn đến giá trị của entropy biến đổi theo từ bit 0 đến bit 1.

Cho một biến ngẫu nhiên rời rạc X , kết quả có thể được sinh ra là x_1, \dots, x_n với xác suất xảy ra là P_{x_1}, \dots, P_{x_n} , entropy của X được thể hiện qua công thức sau:

$$H(X) = - \sum_{i=1}^n P(x_i) \log P(x_i)$$

Ví dụ: Hãy kể cho một người nào đó nghe về thông tin mà họ đã biết rồi, người đó sẽ không học thêm được gì cả. Vì vậy entropy trong trường hợp này sẽ rất thấp.

Nhưng nếu kể cho người đó về những thông tin mà người đó biết rất ít, hoặc chưa biết, thì người đó sẽ có được thông tin mới. Thông tin mới này có giá trị với họ. Người đó học được điều mới. Entropy trong trường hợp này sẽ có giá trị cao.

- Cross-Entropy: cho 2 phân phối p và q , cross entropy đo lượng thông tin trung bình cần thiết để nhận dạng các mẫu ngẫu nhiên từ q khi sử dụng lược đồ mã hoá tối ưu được xây dựng cho p .