다변수 확률 변수의 Entropy

수업 목표

이번 수업의 핵심:

- 두 개 이상의 확률 변수에 대한 Entropy 정의
- Joint entropy의 성질과 이에 대한 증명
- Conditional entropy의 개념과 특성

핵심 개념

- Joint entropy
- Conditional entropy

Joint Entropy

Joint Entropy: 두 개 이상의 확률 변수에 대한 Entropy

- 확률 변수 X, Y에 대해서 X, Y를 모두 알아냄을 통해 얻을 수 있는 정보의 기대값

$$H(X,Y) \coloneqq -\mathbb{E}_{x,y\sim p(X,Y)}[\log p(x,y)]$$

- 이산 확률 변수

$$H(X,Y) = -\sum_i p_{X,Y}(x_i, y_i) \log p_{X,Y}(x_i, y_i)$$

- 연속 확률 변수

$$H(X,Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y) \log p(x,y)\, dxdy$$

Joint Entropy의 성질

- $X = Y$일 때, i.e., $p_{X,Y}(x_i, y_i) = p_X(x_i) = p_Y(y_i)$

$$H(X) = H(Y) = H(X,Y)$$

증명)

$$\begin{aligned} H(X,Y) &= -\sum_i p_{X,Y}(x_i, y_i) \log p_{X,Y}(x_i, y_i) \\ &= -\sum_i p_X(x_i) \log p_X(x_i) = H(X) \\ &= -\sum_i p_Y(y_i) \log p_Y(y_i) = H(Y) \end{aligned}$$

Joint Entropy의 성질

- $X = Y$일 때, i.e., $p_{X,Y}(x_i, y_i) = p_X(x_i) = p_Y(y_i)$

$$H(X) = H(Y) = H(X, Y)$$

- X와 Y가 동일하다는 것은 X와 Y의 값이 항상 동일하게 나온다는 것

1. X와 Y가 같으므로 X를 알아냄을 통해서 얻을 수 있는 정보량과
   Y를 알아냄을 통해서 얻을 수 있는 정보량이 같음
   - $H(X) = H(Y)$
2. X만 알아내도 Y를 자동으로 알 수 있으므로
   X를 알아내어 얻는 정보량과 X, Y를 알아내어 얻는 정보량과 같음
   - $H(X) = H(X, Y)$

Joint Entropy의 성질

- X, Y가 통계적 독립일 때, i.e., $p_{X,Y}(x_i, y_i) = p_X(x_i)\, p_Y(y_i)$

$$H(X) + H(Y) = H(X, Y)$$

통계적 독립

- $p_{X,Y}(x_i, y_i) = p_X(x_i)\, p_Y(y_i)$
- $p_X(x_i|y_i) = p_X(x_i)$

  $\Rightarrow y_i$를 조건으로 주어도 확률 분포가 변하지 않음
- 또는 $p_Y(y_i|x_i) = p_Y(y_i)$

증명)

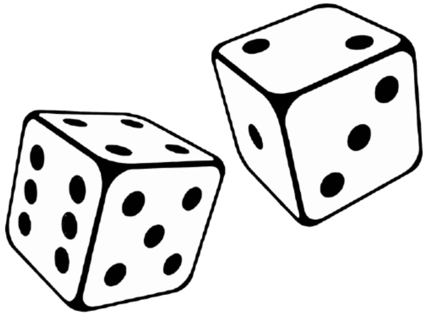
$$\begin{aligned} H(X,Y) &= -\sum_i p_{X,Y}(x_i, y_i) \log p_{X,Y}(x_i, y_i) \\ &= -\sum_i p_X(x_i) p_Y(y_i) \log p_X(x_i) p_Y(y_i) \\ &= -\sum_i p_X(x_i) \log p_X(x_i) - \sum_i p_Y(y_i) \log p_Y(y_i) \\ &= H(X) + H(Y) \end{aligned}$$

Joint Entropy의 성질

- X, Y가 통계적 독립일 때, i.e., $p_{X,Y}(x_i, y_i) = p_X(x_i)\, p_Y(y_i)$

$$H(X) + H(Y) = H(X, Y)$$

- X, Y가 독립이라는 의미는 X가 어떻게 나오던 Y와는 관계가 없다는 것
- X, Y를 모두 알아내어 얻는 정보는 X, Y가 어떻게 나오는지 각각 알아내야 함
  - 주사위와 동전을 각각 던질 때 동전이 앞면이 나왔다고 주사위의 정보를 얻어낼 수 없음
  - 주사위와 동전 모두의 정보를 얻으려면 각각의 정보를 온전히 따로 얻어야 함



Joint Entropy의 성질

일반화

$$H(X), H(Y) \leq H(X, Y) \leq H(X) + H(Y)$$

1. X, Y의 모두 알아냄을 통해 얻는 정보량은
X 또는 Y만 알아냈을 때의 정보량 보다는 크거나 같음
   - $H(X), H(Y) \leq H(X, Y)$

1. X, Y의 모두 알아냄을 통해 얻는 정보량은
X와 Y를 각각 따로 알아내어 얻어지는 정보량의 합보다는 작거나 같음
   - $H(X, Y) \leq H(X) + H(Y)$

Conditional Entropy

Conditional Entropy:

- X를 이미 알고 있을 때 Y를 새로 알려주면 얻는 정보량의 기대값

$$H(Y|X) := -\mathbb{E}_{\underbrace{x,y\sim p(X,Y)}_{\text{주의!}}}[\log p(y|x)]$$

- 이산 확률 변수

$$H(Y|X) = -\sum_i p_{X,Y}(x_i, y_i) \log p_{X,Y}(y_i|x_i)$$

- 연속 확률 변수

$$H(Y|X) = -\int_{-\infty}^{\infty} p(x,y) \log p(y|x)\, dx$$

Conditional Entropy의 특성

- $H(X,Y) = H(Y|X) + H(X) = H(X|Y) + H(Y)$
  - $H(X), H(Y) \leq H(X,Y) \leq H(X) + H(Y)$을 등식으로 변형한 꼴
- X,Y를 모두 알기 위해 필요한 정보량은 X를 알기 위해 필요한 정보량과 X가 주어졌을 때 Y를 알기 위해 필요한 정보량의 합
  - $H(X,Y) = H(X) + H(Y|X)$
  - 또는 반대로 Y를 먼저 알고 Y를 알고 있을 때 X를 알기 위해 필요한 정보량의 합
  - $H(X,Y) = H(Y) + H(X|Y)$

증명)

$$H(Y|X) = -\sum_i p_{X,Y}(x_i, y_i) \log p_{X,Y}(y_i|x_i)$$

$$= -\sum_i p_{X,Y}(x_i, y_i) \log \frac{p_{X,Y}(x_i, y_i)}{p_X(x_i)}$$

$$= H(X,Y) - H(X)$$

요약

- Joint entropy를 이용한 두 개 이상의 확률 변수에 대한 Entropy 정의
- Joint entropy의 여러가지 성질 및 특성
- Conditional entropy를 통한 한 확률 변수를 알았을 때,
  다른 확률 변수를 알기 위한 정보량 계산