AI2100 Deep Learning (HW1)

Saksham Mittal - AI22BTECH11024

Question 4: Joint Entropy

The joint entropy $H(X,Y)$ of a pair of discrete random variables X and Y with joint distribution $p(x,y)$ is defined as:

$$H(X,Y) = -E_{(x,y)\sim p(x,y)}\left[\log p(X,Y)\right]$$

The conditional entropy $H(Y|X)$ is defined as:

$$H(Y|X) = -E_{(x,y)\sim p(x,y)}\left[\log p(Y|X)\right]$$

We need to show that:

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

Explanation

Let's start by expanding the joint entropy $H(X,Y)$:

$$H(X,Y) = -\sum_{x\in X, y\in Y} p(x,y)\log p(x,y)$$

We know that the joint probability $p(x,y)$ can be factored as:

$$p(x,y) = p(x)\cdot p(y|x)$$

Substituting this into the expression for $H(X,Y)$:

$$H(X,Y) = -\sum_{x\in X, y\in Y} p(x,y)\log\left[p(x)\cdot p(y|x)\right]$$

Using the logarithm property $\log(ab) = \log(a)+\log(b)$, we can rewrite the above as:

$$H(X,Y) = -\sum_{x\in X, y\in Y} p(x,y)\left[\log p(x) + \log p(y|x)\right]$$

Distributing the summation:

$$H(X,Y) = -\sum_{x\in X, y\in Y} p(x,y)\log p(x) - \sum_{x\in X, y\in Y} p(x,y)\log p(y|x)$$

The first term simplifies to:

$$-\sum_{x\in X}[p(x)\log p(x)] = H(X)$$

The second term simplifies to:

$$-\sum_{x\in X} p(x)\sum_{y\in Y} p(y|x)\log p(y|x) = H(Y|X)$$

Thus, we have shown that:

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

Question 5: Mutual Information

The mutual information $I(X;Y)$ between two random variables X and Y is defined as:

$$I(X;Y) = E_{(x,y)\sim p(x,y)}\left[\log \frac{p(X,Y)}{p(X)p(Y)}\right]$$

We need to show that:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Explanation

Let's start by considering the mutual information expression:

$$I(X;Y) = E_{(x,y)\sim p(x,y)}\left[\log \frac{p(X,Y)}{p(X)p(Y)}\right]$$

We can rewrite the mutual information using the joint entropy $H(X,Y)$, the marginal entropies $H(X)$ and $H(Y)$, and the conditional entropies $H(X|Y)$ and $H(Y|X)$.

By definition, the joint entropy is:

$$H(X,Y) = -\sum_{x\in X, y\in Y} p(x,y)\log p(x,y)$$

The marginal entropy of X is:

$$H(X) = -\sum_{x\in X} p(x)\log p(x)$$

The conditional entropy $H(X|Y)$ is:

$$H(X|Y) = -\sum_{x\in X, y\in Y} p(x,y)\log p(x|y)$$

The mutual information $I(X;Y)$ can be expressed as:

$$I(X;Y) = H(X) - H(X|Y)$$

Alternatively, it can be written as:

$$I(X;Y) = H(Y) - H(Y|X)$$

Thus, we have shown that the mutual information can be derived from the marginal and conditional entropies of X and Y.

Problem 6: Information Entropy Calculations

Given the joint probabilities:

$$p(X=0,Y=0) = \frac{1}{3}, \quad p(X=0,Y=1) = \frac{1}{3}, \quad p(X=1,Y=1) = \frac{1}{3}, \quad p(X=1,Y=0) = 0$$

1. **Marginal Probabilities:**

$$\begin{aligned}
p(X=0) &= p(X=0,Y=0) + p(X=0,Y=1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \\
p(X=1) &= p(X=1,Y=0) + p(X=1,Y=1) = 0 + \frac{1}{3} = \frac{1}{3} \\
p(Y=0) &= p(X=0,Y=0) + p(X=1,Y=0) = \frac{1}{3} + 0 = \frac{1}{3} \\
p(Y=1) &= p(X=0,Y=1) + p(X=1,Y=1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}
\end{aligned}$$

2. **Entropy of X:**

$$\begin{aligned}
H(X) &= -\sum_x p(x) \log_2 p(x) \\
&= -\left[\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3}\right] \\
&= -\left[\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3}\right] \\
&\approx 0.918 \text{ bits}
\end{aligned}$$

3. **Entropy of Y:**

$$\begin{aligned}
H(Y) &= -\sum_y p(y) \log_2 p(y) \\
&= -\left[\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3}\right] \\
&\approx 0.918 \text{ bits}
\end{aligned}$$

4. **Joint Entropy** $H(X,Y)$**:**

$$\begin{aligned}
H(X,Y) &= -\sum_{x,y} p(x,y)\log_2 p(x,y) \\
&= -\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{1}{3}\log_2\frac{1}{3} + \frac{1}{3}\log_2\frac{1}{3}\right] \\
&= -\left[3 \times \frac{1}{3}\log_2\frac{1}{3}\right] \\
&\approx 1.585 \text{ bits}
\end{aligned}$$

5. **Conditional Entropy** $H(X|Y)$**:**

$$\begin{aligned}
H(X|Y) &= H(X,Y) - H(Y) \\
&= 1.585 - 0.918 \\
&\approx \frac{2}{3} \text{ bits} \\
&= 0.667 \text{ bits}
\end{aligned}$$

6. **Conditional Entropy** $H(Y|X)$**:**

$$\begin{aligned}
H(Y|X) &= H(X,Y) - H(X) \\
&= 1.585 - 0.918 \\
&\approx \frac{2}{3} \text{ bits} \\
&= 0.667 \text{ bits}
\end{aligned}$$

7. **Mutual Information** $I(X;Y)$**:**

$$\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= 0.918 - 0.667 \\
&\approx 0.251 \text{ bits}
\end{aligned}$$