February 27, 2021

2.12 Expressing mutual information in terms of entropies

Show that $\mathbb{I}(X, Y) = H(X) - H(X|Y)$

$$
\begin{aligned}
\mathbb{I}(X, Y) &= \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\
&= \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\
&= \sum_x \sum_y p(x, y) \log \frac{p(x|y)}{p(x)} \\
&= \sum_x \sum_y p(x, y) \log p(x|y) - \sum_x \sum_y p(x, y) \log p(x) \\
&= \sum_y \sum_x p(x|y)p(y) \log p(x|y) - \sum_x \log p(x) \sum_y p(x, y) \\
&= \sum_y p(y) \sum_x p(x|y) \log p(x|y) - \sum_x p(x) \log p(x) \\
&= -H(X|Y) + H(X)
\end{aligned}
$$