

- For each input distribution $p(x)$, we have

$$p(x, y) = p(x)p(y|x).$$

- From $p(x, y)$, we can compute $I(X; Y)$.
- Maximize $I(X; Y)$ over all $p(x)$.