Mutual Information and its Various Invariances

Tyler Feemster

June 25, 2024

The mutual information of two continous, real-valued random variables X and Y is defined by

$$I(X;Y) = \int_{\mathbb{R}} \int_{\mathbb{R}} P_{(X,Y)}(x,y) \log\left(\frac{P_{(X,Y)}(x,y)}{P_X(x)P_Y(y)}\right) dxdy$$

1 Invertible, Differentiable Functions

Let $f : \mathbb{R} \to \mathbb{R}$ be an invertible, differentiable function, and let $X' := f(X)$.

Theorem 1. $I(X;Y) = I(X';Y)$.

Proof. First, note that

$$P_X(x)dx = P_{X'}(f(x)) \cdot df(x) = P_{X'}(f(x)) \cdot f'(x)dx,$$

and similarly,

$$P_{(X,Y)}(x,y)dx = P_{(X',Y)}(f(x),y) \cdot f'(x)dx.$$

Therefore, we have

$$\begin{aligned} I(X;Y) &= \int_{\mathbb{R}} \int_{\mathbb{R}} P_{(X,Y)}(x,y) \log\left(\frac{P_{(X,Y)}(x,y)}{P_X(x)P_Y(y)}\right) dxdy \\ &= \int\int P_{(X',Y)}(f(x),y)f'(x) \log\left(\frac{P_{(X',Y)}(f(x),y)f'(x)}{P_{X'}(f(x))f'(x)P_Y(y)}\right) dxdy \\ &= \int\int P_{(X',Y)}(f(x),y) \log\left(\frac{P_{(X',Y)}(f(x),y)}{P_{X'}(f(x))P_Y(y)}\right) f'(x)dxdy \\ &= \int\int P_{(X',Y)}(x',y) \log\left(\frac{P_{(X',Y)}(x',y)}{P_{X'}(x')P_Y(y)}\right) dx'dy = I(X';Y). \end{aligned}$$

□

Mutual information is invariant under invertible, continuously differentiable transformations.

Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be invertible, continuously differentiable transformations, and define $X' := f(X)$ and $Y' := g(Y)$. By symmetry and Theorem 1, we have

$$I(X;Y) = I(X';Y) = I(X';Y').$$

2 Rotations

Because mutual information is independent with respect to continuously differentiable, invertible transformations, we now turn to the question of "mixing" X and Y. The most natural transformations that achieve this are rotations in the XY plane. Rotating by θ, the differentials become:

$$dx = \cos\theta dx' + \sin\theta dy'$$
$$dy = -\sin\theta dx' + \cos\theta dy'$$
$$dx \wedge dy = \cos^2\theta dx' \wedge dy' - \sin^2\theta dy' \wedge dx' = (\cos^2\theta + \sin^2\theta) dx' \wedge dy' = dx' \wedge dy'$$

2.1 Gaussians

Because mutual information is invariant under invertible transformations, we can coerce P_X and P_Y to follow a standard normal distribution without affecting their mutual information. So, assume

$$P_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad \text{and} \qquad P_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

Theorem 2. *If X and Y are standard normal distributions and X' and Y' are the random variables after a rotation, $I(X;Y) = I(X';Y')$.*

Proof. We have

$$I(X;Y) = \int\int P_{(X,Y)}(x,y) \log\left(\frac{P_{(X,Y)}(x,y)}{P_X(x)P_Y(y)}\right) dxdy$$
$$= \int\int P_{(X,Y)}(x,y) \log\left(\frac{P_{(X,Y)}(x,y)}{\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}}\right) dxdy$$
$$= \int\int P_{(X,Y)}(x,y) \log\left(\frac{2\pi P_{(X,Y)}(x,y)}{e^{-(x^2+y^2)/2}}\right) dxdy$$
$$= \int\int P_{(X,Y)}(x,y) \left(\log 2\pi + \log P_{(X,Y)}(x,y) + \frac{x^2+y^2}{2}\right) dxdy$$
$$= \log 2\pi + \int\int P_{(X,Y)}(x,y) \left(\log P_{(X,Y)}(x,y) + \frac{x^2+y^2}{2}\right) dxdy.$$

Now, note that under a rotation by θ,

$$dxdy = dx'dy', \qquad P_{(X,Y)}(x,y) = P_{(X',Y')}(x',y'), \text{ and } \qquad x^2+y^2 = x'^2+y'^2.$$

Therefore,

$$I(X;Y) = \log 2\pi + \int\int P_{(X,Y)}(x,y) \left(\log P_{(X,Y)}(x,y) + \frac{x^2+y^2}{2}\right) dxdy$$
$$= \log 2\pi + \int\int P_{(X',Y')}(x',y') \left(\log P_{(X',Y')}(x',y') + \frac{x'^2+y'^2}{2}\right) dx'dy' = I(X';Y'),$$

as desired. $\square$

Mutual Information of two normally distributed variables (marginally) is invariant under rotation.

2.2 Mutual Information

When X and Y are standard normal,

$$H(X) = H(Y) = \frac{1}{2}(1 + \log 2\pi).$$

Therefore, in the case when the marginal distributions of $P_{(X,Y)}$ are standard normal,

$$I(X;Y) = H(X,Y) - H(X) - H(Y) = H(X,Y) - (1 + \log 2\pi).$$

2.3 More General Distributions