Information Theory

Tyler Feemster

June 25, 2024

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

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

Invariance Properties

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

Lemma 1. $H(X) - H(X,Y) = H(X') - H(X',Y)$.

Proof. We start by expanding the left-hand side and substituting $x \mapsto f(x)$. First, note that

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

Then,

$$\begin{aligned}
H(X) - H(X,Y) &= -\int_{\mathbb{R}} P_X(x) \log P_X(x) dx + \int_{\mathbb{R}} \int_{\mathbb{R}} P_{(X,Y)}(x,y) \log P_{(X,Y)}(x,y) dxdy \\
&= -\int P_{X'}(f(x)) f'(x) \log(P_{X'}(f(x)) f'(x)) dx \\
&\quad + \int \int P_{(X',Y)}(f(x),y) f'(x) \log(P_{(X',Y)}(f(x),y) f'(x)) dxdy \\
&= -\int P_{X'}(f(x)) \log P_{X'}(f(x)) f'(x) dx - \int P_{X'}(f(x)) f'(x) \log f'(x) dx \\
&\quad + \int \int P_{(X',Y)}(f(x),y) \log P_{(X',Y)}(f(x),y) f'(x) dxdy + \int \int P_{(X',Y)}(f(x),y) f'(x) \log f'(x) dxdy \\
&= -\int P_{X'}(x') \log P_{X'}(x') dx' + \int \int P_{(X',Y)}(x',y) \log P_{(X',Y)}(x',y) dx'dy \\
&\quad - \int P_{X'}(f(x)) f'(x) \log f'(x) dx + \int \int P_{(X',Y)}(f(x),y) f'(x) \log f'(x) dxdy \\
&= H(X') - H(X',Y) - \int P_{X'}(f(x)) f'(x) \log f'(x) dx + \int \int P_{(X',Y)}(f(x),y) f'(x) \log f'(x) dxdy.
\end{aligned}$$

We complete the lemma by observing

$$\int\int P_{(X',Y)}(f(x),y)f'(x)\log f'(x)dxdy = \int\left(\int P_{(X',Y)}(f(x),y)dy\right)f'(x)\log f'(x)dx$$
$$= \int P_{X'}(f(x))f'(x)\log f'(x)dx.$$

$\square$

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

Proof. This follows immediately from the identity

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

and Lemma 1. $\square$

Mutual information is invariant under invertible, continuously differentiable transformations.

Root Jensen-Shannon Divergence (rJSD)

$$JS(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M),$$

where $M = \frac{1}{2}(P+Q)$ and

$$D_{KL}(P||Q) = \int P(x)\log\frac{P(x)}{Q(x)}dx$$

$$P(x)dx = P(f(x))df(x) = P(f(x))f'(x)dx,$$

so

$$P(x) = P'(f(x))f'(x) \qquad \text{and} \qquad Q(x) = Q'(f(x))f'(x)$$

$$M(x) = \frac{1}{2}(P(x)+Q(x)) = \frac{f'(x)}{2}(P'(f(x)) + Q'(f(x)))$$

$$\begin{aligned}
D_{KL}(P||M) &= \int P(x)\log\frac{P(x)}{M(x)}dx \\
&= \int P'(f(x))f'(x)\log\frac{P'(f(x))f'(x)}{\frac{f'(x)}{2}(P'(f(x))+Q'(f(x)))}dx \\
&= \int P'(f(x))\log\frac{P'(f(x))}{\frac{1}{2}(P'(f(x))+Q'(f(x)))}df(x) \\
&= \int P'(x)\log\frac{P'(x)}{\frac{1}{2}(P'(x)+Q'(x))}dx \\
&= \int P'(x)\log\frac{P'(x)}{M'(x)}dx = D_{KL}(P'||M')
\end{aligned}$$

Mutual Information

Because mutual information is independent with respect to continuously differentiable, bijective transformations, it's interesting to consider the effect of rotations in the XY plane.

We have

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

$$\begin{aligned}
dx &= \cos\theta dx' + \sin\theta dy' \\
dy &= \cos\theta dy' - \sin\theta dx' \\
dx\wedge dy &= \cos^2\theta dx'\wedge dy' - \sin^2\theta dy'\wedge dx' = dx'\wedge dy'
\end{aligned}$$

$$\begin{aligned}
P_X(x)dx + P_Y(y)dy &= P_{X'}(x')dx' + P_{Y'}(y')dy' \\
P_X(x)(\cos\theta dx' + \sin\theta dy') + P_Y(y)(\cos\theta dy' - \sin\theta dx') &= P_{X'}(x')dx' + P_{Y'}(y')dy' \\
P_X(x)\cos\theta dx' + P_X(x)\sin\theta dy' + P_Y(y)\cos\theta dy' - P_Y(y)\sin\theta dx' &= P_{X'}(x')dx' + P_{Y'}(y')dy'
\end{aligned}$$

Gaussian

Because mutual information is invariant under invertible transformations, we can coerce P_X and P_Y to follow a normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Then,

$$\begin{aligned}
I(X;Y) &= \int_{\mathcal{X}\times\mathcal{Y}} P_{(X,Y)}(x,y) \log\left(\frac{P_{(X,Y)}(x,y)}{P_X(x)P_Y(y)}\right) dxdy \\
&= \int_{\mathcal{X}\times\mathcal{Y}} P_{(X,Y)}(x,y) \log\left(\frac{P_{(X,Y)}(x,y)}{\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}}\right) dxdy \\
&= \int_{\mathcal{X}\times\mathcal{Y}} P_{(X,Y)}(x,y) \log\left(\frac{2\pi P_{(X,Y)}(x,y)}{e^{-(x^2+y^2)/2}}\right) dxdy \\
&= \int_{\mathcal{X}\times\mathcal{Y}} P_{(X,Y)}(x,y) \left(\log 2\pi + \log P_{(X,Y)}(x,y) + \frac{x^2+y^2}{2}\right) dxdy \\
&= \log 2\pi + \int_{\mathcal{X}\times\mathcal{Y}} P_{(X,Y)}(x,y) \left(\log P_{(X,Y)}(x,y) + \frac{x^2+y^2}{2}\right) dxdy
\end{aligned}$$

Now, if we rotate X and Y:

$$\begin{aligned}
dx &= \cos\theta dx' + \sin\theta dy' \\
dy &= \cos\theta dy' - \sin\theta dx' \\
dx \wedge dy &= dx' \wedge dy'
\end{aligned}$$

Then,

$$\begin{aligned}
I(X;Y) &= \log 2\pi + \int_{\mathcal{X}\times\mathcal{Y}} P_{(X,Y)}(x,y) \left(\log P_{(X,Y)}(x,y) + \frac{x^2+y^2}{2}\right) dxdy \\
&= \log 2\pi + \int_{\mathcal{X}'\times\mathcal{Y}'} P_{(X',Y')}(x',y') \left(\log P_{(X',Y')}(x',y') + \frac{x'^2+y'^2}{2}\right) dx'dy' = I(X';Y'),
\end{aligned}$$

so

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

Gaussian Shift

Let f be the invertible transformation such that $P \circ f = X$, where X is a normal distribution. P is a distribution over $\mathbb{R}$. Then,

$$\int_{-\infty}^{x} P(f(t))f'(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}dt$$
$$\sqrt{2\pi}P(f(x))f'(x) = e^{-x^2/2}$$
$$-x^2/2 = \frac{1}{2}\log 2\pi + \log P(f(x)) + \log f'(x)$$
$$x^2 = -\log 2\pi - 2\log P(f(x)) - 2\log f'(x)$$

If g is the inverse of f,

$$g(x)^2 = -\log 2\pi - 2\log P(x) - 2\log f'(g(x)),$$

but since $g'(x) = \frac{1}{f'(g(x))}$ (definition of inverse and chain rule), we have

$$g(x)^2 = \log\left(\frac{g'(x)^2}{2\pi P(x)^2}\right).$$

Also,

$$e^{g(x)^2/2} = \frac{g'(x)}{\sqrt{2\pi}P(x)}$$
$$g(x)g'(x)e^{g(x)^2/2} = \frac{1}{\sqrt{2\pi}} \cdot \frac{P(X)g''(x) - P'(x)g'(x)}{P(x)^2}$$
$$g(x)g'(x)\frac{g'(x)}{\sqrt{2\pi}P(x)} = \frac{1}{\sqrt{2\pi}} \cdot \frac{P(x)g''(x) - P'(x)g'(x)}{P(x)^2}$$
$$g(x)g'(x)^2P(x) = P(x)g''(x) - P'(x)g'(x)$$
$$[g(x)g'(x)P(x) + P'(x)]g'(x) = P(x)g''(x)$$
$$g(x)g'(x) + \frac{P'(x)}{P(x)} = \frac{g''(x)}{g'(x)}$$
$$g(x)g'(x) + \frac{d}{dx}\log P(x) = \frac{d}{dx}\log g'(x)$$