Machine Learning HW2

21300126 Seungwoo Kim

<Bishop>

1.32 Consider a vector $\mathbf{x}$ of continuous variables with distribution $p(\mathbf{x})$ and corresponding entropy $\mathrm{H}[\mathbf{x}]$. Suppose that we make a nonsingular linear transformation of $\mathbf{x}$ to obtain a new variable $\mathbf{y} = \mathrm{A}\mathbf{x}$. Show that the corresponding entropy is given by $\mathrm{H}[\mathbf{y}] = \mathrm{H}[\mathbf{x}] + \ln|\mathrm{A}|$ where $|\mathrm{A}|$ denotes the determinant of A.

$\mathbf{y} = \mathrm{A}\mathbf{x}$ can be rewritten as $\mathbf{x} = \frac{1}{\mathrm{A}}\boldsymbol{y}$, and we can derive $p(y) = p(x)\left|\frac{dx}{dy}\right| = p(x)\left|\frac{1}{A}\right|$.

$$H[y] = -\int p(y)\ln p(y)\,dy = -\int \frac{p(x)}{|A|}\ln\frac{p(x)}{|A|}\left|\frac{dy}{dx}\right|dx - \int p(x)\ln\frac{p(x)}{|A|}\,dx$$

$$= -\int p(x)\ln p(x)dx + \ln|A|\int p(x)dx = H[x] + \ln|A|$$

This equation can be derived due to the fact that $\int p(x)dx = 1$. Therefore, we can derive entropy of $\mathbf{y} = \mathrm{A}\mathbf{x}$ as $\mathrm{H}[\mathbf{y}] = \mathrm{H}[\mathbf{x}] + \ln|\mathrm{A}|$.

1.35 Use the results (1.106) and (1.107) to show that the entropy of the univariate Gaussian (1.109) is given by (1.110).

Probability of Gaussian distribution is $p(x) = \frac{1}{(2\pi\sigma^2)^{1/2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, and the entropy is calculated using this probability.

$$H[x] = -\int p(x)\ln p(x)\,dx = -\int p(x)\ln \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{(2\pi\sigma^2)^{\frac{1}{2}}}dx$$

$$= \int p(x)\frac{(x-\mu)^2}{2\sigma^2}dx + \int p(x)\ln(2\pi\sigma^2)^{\frac{1}{2}}\,dx$$

$$= \frac{1}{2\sigma^2}\int p(x)(x-\mu)^2 dx + \frac{1}{2}\int p(x)\ln(2\pi\sigma^2)\,dx$$

Using the property that $\int p(x)\,dx = 1$ and $\int p(x)(x-\mu)^2 dx = \sigma^2$,

$$H[x] = \frac{\sigma^2}{2\sigma^2} + \frac{1}{2}\ln(2\pi\sigma^2) = \frac{1}{2}(1 + \ln(2\pi\sigma^2))$$

, which is the entropy of univariate Gaussian.

1.37. Using the definition (1.111) together with the product rule of probability, prove the result (1.112).

Using the property of conditional probability, conditional entropy of y given x can be divided into as following;

$$\mathrm{H}[y|x] = -\int\int p(x,y)\ln p(y|x)\ dydx = -\int\int p(x,y)\ln\frac{p(x,y)}{p(x)}\ dydx$$

$$= -\int\int p(x,y)\ln p(x,y)\,dydx + \int\int p(x,y)\ln p(x)\,dydx$$

$$= \mathrm{H}[x,y] - \int p(x)\ln p(x)\,dx = \mathrm{H}[x,y] - \mathrm{H}[x]$$

Therefore, $\mathrm{H}[x,y] = \mathrm{H}[y|x] + \mathrm{H}[x]$.

1.40 By applying Jensen's inequality (1.115) with $f(x) = \ln x$, show that the arithmetic mean of a set of real numbers is never less than their geometrical mean.

According to Jensen's Inequality, when the function $f(x)$ is a convex function, then following inequality is satisfied:

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \leq \sum_{i=1}^{M} \lambda_i f(x_i)$$

However, the function $f(x) = \ln x$ is a concave function, so there are two ways to deal with Jensen's Inequality: either negate the function itself or change the direction of inequality in reverse way.

Just following the second way, the inequality looks like

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \geq \sum_{i=1}^{M} \lambda_i f(x_i)$$

, and the fact that $\sum_{i=1}^{M} \lambda_i = 1$, let's consider $\lambda_i = \frac{1}{M}$. Therefore,

$$f\left(\frac{1}{M}\sum_{i=1}^{M} x_i\right) \geq \frac{1}{M}\sum_{i=1}^{M} f(x_i)$$

$$\frac{1}{M}\ln \sum_{i=1}^{M} x_i \geq \frac{1}{M}\sum_{i=1}^{M} \ln x_i$$

$$\frac{1}{M}\ln(x_1 + x_2 + \cdots + x_M) \ \geq \frac{1}{M}(\ln x_1 + \ln x_2 + \cdots + \ln x_M)$$

$$(\ln x_1 \times \ln x_2 \times \cdots \times \ln x_M)^{\frac{1}{\mathrm{M}}} \ \geq \frac{1}{M}(\ln x_1 + \ln x_2 + \cdots + \ln x_M)$$

The left-hand side is the geometric mean, and the right-hand side is the arithmetic mean. Therefore, it is proven that the arithmetic mean of $f(x) = \ln x$ is never less than the geometric mean of it.

1.41 Using the sum and product rules of probability, show that the mutual information $I(x, y)$ satisfies the relation (1.121).

$$I[x,y] = -\int\int p(x,y)\ln\left(\frac{p(x)p(y)}{p(x,y)}\right)dxdy$$

$$= -\int\int p(x,y)\ln p(x)\, dydx \;\; -\int\int p(x,y)\ln\left(\frac{p(y)}{p(x,y)}\right)dxdy$$

$$= -\int p(x)\ln p(x)\, dx \;\; +\int\int p(x,y)\ln\left(\frac{p(x,y)}{p(y)}\right)dxdy$$

$$= H[x] + \int\int p(x,y)\ln p(x|y)\, dydx = H[x] - H[x|y]$$

Similarly, when we first take p(y) out from the equation, then

$$I[x,y] = -\int\int p(x,y)\ln\left(\frac{p(x)p(y)}{p(x,y)}\right)dxdy$$

$$= -\int\int p(x,y)\ln p(y)\, dxdy \;\; -\int\int p(x,y)\ln\left(\frac{p(x)}{p(x,y)}\right)dxdy$$

$$= -\int p(y)\ln p(y)\, dy \;\; +\int\int p(x,y)\ln\left(\frac{p(x,y)}{p(x)}\right)dxdy$$

$$= H[y] + \int\int p(x,y)\ln p(y|x)\, dxdy = H[y] - H[y|x]$$

Therefore, $I[x,y] = H[x] - H[x|y] = H[y] - H[y|x]$.