

ECE 612, Information Theory

Ganyu (Bruce) Xu (g66xu)

Winter, 2024

Preliminaries

Definition 0.1. The normal distribution $N(\mu, \sigma^2)$ has the probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

Definition 0.2. The joint normal distribution $N(\mu, K)$ is defined by probability density function:

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(K)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top K^{-1}(\mathbf{x} - \mu)\right)$$

Theorem 0.1 (Joint normality implies marginal normality). If $\mathbf{X} = (X_1, X_2, \dots, X_n)$ follows a joint normal distribution, then any linear combination of \mathbf{X} follows normal distribution.

1 Entropy, mutual information, divergence

Theorem 1.1. Let X, Y be random variables. $I(X; Y)$ is concave with respect to the probability distribution of X . For a fixed marginal distribution of X , $I(X; Y)$ is convex with respect to $f_{Y|X}$.

Theorem 1.2 (Fano's inequality). Let $X \rightarrow Y \rightarrow \hat{X}$ represent an encode-decode process, where $X, \hat{X} \in \mathcal{X}$ have the same support. Let e denote decoding error $\hat{X} \neq X$, then:

$$H(X | Y) \leq H(P_e) + P_e \log(|\mathcal{X}|)$$

2 Entropy rate

3 Asymptotic equipartition property

4 Data compressions

5 Channel capacity

6 Differential entropy

Definition 6.1. Let $X \in \mathbb{R}$ be a random variable with probability density function f_X . The **differential entropy** of X is defined by

$$h(X) = - \int_{x \in \mathbb{R}} f_X(x) \ln(f_X(x)) dx = -E[\ln(f_X(X))]$$

Some notable results:

- For uniform distribution over $[0, a]$, $f_X(x) = \frac{1}{a}$, $h(X) = \ln(a)$
- For $X \sim N(0, \sigma^2)$, $h(X) = \frac{1}{2} \ln(2\pi e \sigma^2)$
- For multivariate normal $\mathbf{X} \sim N(\mathbf{0}, K)$, $h(\mathbf{X}) = \frac{1}{2} \log((2\pi e)^n \det K)$
- For some constant a , $h(X + a) = h(X)$
- For some constant a , $h(aX) = \ln(|a|) + h(X)$

Theorem 6.1. *Let X be continuous random variable with 0 mean and σ^2 variance, then*

$$h(X) \leq \frac{1}{2} \ln(2\pi e \sigma^2)$$

Equality is reached if and only if X is Gaussian

7 Gaussian channel

Definition 7.1 (Information channel capacity). *Let $Y = X + Z$, where $Z \stackrel{s}{\leftarrow} N(0, \sigma^2)$ and $E[X^2] \leq P$ for some power level constraint P . The **information channel capacity** is defined by*

$$C^I = \max_{f_X: E[X^2] \leq P} I(X; Y)$$

Theorem 7.1. *The information channel capacity of a Gaussian channel is*

$$\max_{f_X: E[X^2] \leq P} I(X; Y) = \frac{1}{2} \log \left(1 + \frac{P}{\sigma^2} \right) \quad (1)$$

Where P is the power constraint, and σ^2 is the variance of the Gaussian noise. The maximum is achieved when X follows Gaussian distribution $X \stackrel{s}{\leftarrow} N(0, P)$

8 Rate distortion theory

Theorem 8.1. *Let X follow Bernoulli(p), then*

$$R(D) = \begin{cases} h(p) - h(D) & \text{When } D < \min(p, 1-p) \\ 0 & \text{otherwise} \end{cases}$$

Theorem 8.2. *For $X \stackrel{s}{\leftarrow} N(0, \sigma^2)$:*

$$R(D) = \begin{cases} \frac{1}{2} \log \left(\frac{\sigma^2}{D} \right) & (D \leq \sigma^2) \\ 0 & \text{otherwise} \end{cases}$$