NCTU Machine Learning Hw2

Liang Yu Pan 0486016

November 26, 2015

1 Information Theory

(a)

Overview We use the differential entropy theorem with **Lagrange multipliers** to prove that the probability distribution which maximize the differential entropy is **Gaussian distribution**.

Useful formula

$$H(x) = -\int p(x) \ln p(x) dx \tag{1}$$

$$\int_{-\infty}^{\infty} p(x) dx = 1 \tag{2}$$

$$\int_{-\infty}^{\infty} x p(x) dx = \mu \tag{3}$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx = \sigma^2 \tag{4}$$

Solve p(x) By **Lagrange multipliers** and formula(1),(2),(3),(4), we get the formula

$$L = -\int_{-\infty}^{\infty} p(x) \ln p(x) dx + \lambda_1 (\int_{-\infty}^{\infty} p(x) dx - 1) + \lambda_2 (\int_{-\infty}^{\infty} x p(x) dx - \mu) + \lambda_3 (\int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx - \sigma^2) \tag{5}$$

Then solve $\frac{\delta L}{\delta p(x)} = 0$, $\frac{\delta L}{\delta p(x)} = -\ln p(x) - 1 + \lambda_1 + \lambda_2 x + \lambda_3 (x-\mu)^2$
We get

$$p(x) = e^{-1+\lambda_1+\lambda_2 x+\lambda_3 (x-\mu)^2} \tag{6}$$

Substitute (6) into (2), (3), (4), we can get

$$\lambda_1 = 1 - \frac{1}{2}\ln 2\pi\sigma^2 \tag{7}$$

$$\lambda_2 = 0 \tag{8}$$

$$\lambda_3 = \frac{1}{2\sigma^2} \tag{9}$$

Then substitute (7),(8),(9) into (5), we can get

$$p(x) = \frac{1}{2\pi\sigma^2}e\{-\frac{(x-\mu)^2}{2\sigma^2}\} \tag{10}$$

which is **Gaussian distribution**.

(b)

derive entropy Substitute (10) into (1), we get $\frac{1}{2}(\ln 2\pi\sigma^2 + 1)$