Individual Part for Assignment 1 of CSE253

Guanghao Chen, A53276390
Department of Computer Science
9500 Gilman Drive, La Jolla, CA
`guc001@eng.ucsd.edu`

1 Problems from Bishop

1.1 Exercise 1.1

Using Equation (1.41) when $\lambda = 2$, the LHS of equation (1.42) can be reduced as

$$\int_{-\infty}^{\infty} e^{-x_i^2} dx_i = \pi^{\frac{1}{2}}$$

$$\prod_{i=1}^{d} \int_{-\infty}^{\infty} e^{-x_i^2} dx_i = \pi^{\frac{d}{2}}$$

(1)

For the RHS, by enforcing $u = r^2$

$$s_d \int_0^{\infty} e^{-r^2} r^{d-1} dr = s_d \int_0^{\infty} u^{\frac{d}{2}-1} e^{-u} d\sqrt{u}$$

$$= \frac{s_d}{2} \int_0^{\infty} u^{\frac{d}{2}-1} e^{-u} du$$

(2)

$$= \frac{s_d}{2} \Gamma(\frac{d}{2})$$

Since LHS should be equal to RHS,

$$\pi^{\frac{d}{2}} = \frac{s_d}{2} \Gamma(\frac{d}{2})$$

$$s_d = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}$$

(3)

For $d = 2$,

$$S_2 = \frac{2\pi^{2/2}}{\Gamma(\frac{2}{2})}$$

$$= \frac{2\pi}{\Gamma(1)}$$

(4)

$$= 2\pi$$

For $d = 3$,

$$S_3 = \frac{\pi^{3/2}}{\Gamma(\frac{3}{2})}$$

$$= \frac{2\pi^{\frac{3}{2}}}{\frac{\sqrt{\pi}}{2}}$$

(5)

$$= 4\pi$$

The results reduces to the common expression of circle and sphere when $r = 1$.

1.2 Exercise 1.2

Since the relationship between Volume V and Surface S for a sphere with arbitrary radius a is

$$\frac{dV}{da} = S$$

Therefore the volume of a hypersphere of radius a in d-dimension should be

$$V_d = \int S da = \int S_d a^{d-1} da = \frac{s_d}{d} a^d \tag{6}$$

Therefore, the volume of hypersphere over the volume of cube will be represented as

$$\frac{V_d}{V_c} = \frac{\frac{S_d a^d}{d}}{(2a)^d} = \frac{S_d}{2^d d} = \frac{2\pi^{d/2}}{2^d d \Gamma(d/2)} = \frac{\pi^{d/2}}{d 2^{d-1} \Gamma(d/2)} \tag{7}$$

Further, we can transform it into the following format, which is

$$\frac{V_d}{V_c} = \frac{\pi^{d/2}}{d 2^{d-1} \Gamma(d/2)} = \frac{d}{2}\left(\frac{\pi}{4}\right)^{\frac{d}{2}} \frac{1}{\Gamma(\frac{d}{2})} \tag{8}$$

When $d \to \infty$, it's easy to find that all the terms in equation (6) converges to zero, so the whole expression converges to zero. In addition, the ratio of distance from the center to one corner over the distance to one side can be represented in the following,

$$\frac{center\ to\ one\ corner}{center\ to\ one\ side} = \frac{\sqrt{d \times a^2}}{a} = \sqrt{d} \tag{9}$$

When d goes larger to ∞, the ratio goes to ∞.

1.3 Exercise 1.3

According to the definition of fraction f, it can be represented as following

$$f = 1 - \frac{V_{a-\epsilon}}{V_a}$$
$$= 1 - \frac{(a - \epsilon)^d}{a^d} \tag{10}$$
$$= 1 - (1 - \frac{\epsilon}{a})^d$$

And further, when $d \to \infty$

$$\lim_{d \to \infty} f = \lim_{d \to \infty} 1 - (1 - \frac{\epsilon}{a})^d$$
$$= 1 - 0 \tag{11}$$
$$= 1$$

Evaluations:
Given $\epsilon/a = 0.01$,

$$f_{d=2} = 1 - 0.99^2 = 0.0199$$
$$f_{d=10} = 1 - 0.99^{10} = 0.0956 \tag{12}$$
$$f_{d=1000} = 1 - 0.99^{1000} = 0.99996$$

When evaluating the fraction lying inside the radius a/2, it means $\epsilon = \frac{a}{2}$ either, which means $\frac{\epsilon}{a} = \frac{1}{2}$

$$1 - f_{d=2} = 0.5^2 = 0.25$$
$$1 - f_{d=10} = 0.5^{10} = 0.000977 \tag{13}$$
$$1 - f_{d=1000} = 0.5^{1000} \approx 0.00000$$

1.4 Exercise 1.4

Given the probabilistic density function, the probabilistic mass can be deduced by computing the integral of $p(x)$ over x in the thin shell. Besides, since the shell is so tiny that we can consider the density as a constant.

$$
\begin{aligned}
\int_{shell} p(x)dx &= p(x)\int_{shell} dx \\
&= p(x)V_{shell} \\
&= \frac{1}{(2\pi\sigma^2)^{1/2}}exp(-\frac{\|x\|^2}{2\sigma^2})s_d r^{d-1}\epsilon \\
&= \frac{s_d r^{d-1}}{(2\pi\sigma^2)^{1/2}}exp(-\frac{\|x\|^2}{2\sigma^2})\epsilon \\
&= \frac{s_d r^{d-1}}{(2\pi\sigma^2)^{1/2}}exp(-\frac{r^2}{2\sigma^2})\epsilon
\end{aligned}
\tag{14}
$$

Since we want to represent the equation into format $\rho(r)\epsilon$, so $\rho(r)$ would be

$$
\frac{s_d r^{d-1}}{(2\pi\sigma^2)^{1/2}}exp(-\frac{r^2}{2\sigma^2})
\tag{15}
$$

Further, we can compute the derivative of equation (13) with respect to variable r

$$
\begin{aligned}
\frac{d\rho(r)}{dr} &= (d-1)\frac{s_d}{(2\pi\sigma^2)^{\frac{d}{2}}}r^{d-2}exp(-\frac{r^2}{2\sigma^2}) - (\frac{r}{\sigma^2}exp(-\frac{r^2}{2\sigma^2})\frac{s_d r^{d-1}}{(2\pi\sigma^2)^{\frac{d}{2}}}) \\
&= \frac{s_d r^{d-2}}{(2\pi\sigma^2)^{\frac{d}{2}}}(d-1-\frac{r^2}{\sigma^2})exp(-\frac{r^2}{2\sigma^2})
\end{aligned}
\tag{16}
$$

Enforcing equation (16) to be 0, we can obtain the root

$$
\hat{r} = \sqrt{d-1}\sigma
\tag{17}
$$

When r is greater than $\hat{r}$, the derivative will be greater than 0 and when r is smaller than $\hat{r}$, the derivative will be smaller than 0, therefore, $\hat{r}$ will be the only maximum point. Since dimension d is much larger than 1, therefore $\hat{r} = \sqrt{d}\sigma$ when d is extremely large.

According to equation (1.50), we can deduce the equation in the following

$$
\begin{aligned}
\frac{\rho(\hat{r}+\epsilon)}{\rho(\hat{r})} &= \frac{\frac{s_d(\hat{r}+\epsilon)^{d-1}}{(2\pi\sigma^2)^{1/2}}exp(-\frac{(\hat{r}+\epsilon)^2}{2\sigma^2})}{\frac{s_d r^{d-1}}{(2\pi\sigma^2)^{1/2}}exp(-\frac{r^2}{2\sigma^2})} \\
&= \frac{(\hat{r}+\epsilon)^{d-1}}{\hat{r}^{d-1}}\frac{exp(-\frac{(\hat{r}+\epsilon)^2}{2\sigma^2})}{exp(-\frac{\hat{r}^2}{2\sigma^2})} \\
&= (1+\frac{\epsilon}{\hat{r}})^{d-1}exp(-\frac{2\epsilon\hat{r}+\epsilon^2}{2\sigma^2}) \\
&= exp(-\frac{2\epsilon\hat{r}+\epsilon^2}{2\sigma^2}+(d-1)ln(1+\frac{\epsilon}{\hat{r}}))
\end{aligned}
\tag{18}
$$

Then expanding the term inside eponential by using Talyor expandation

$$
\begin{aligned}
-\frac{2\epsilon\hat{r}+\epsilon^2}{2\sigma^2}+(d-1)ln(1+\frac{\epsilon}{\hat{r}}) &\approx -\frac{2\epsilon\hat{r}+\epsilon^2}{2\sigma^2}+(d-1)(\frac{\epsilon}{\hat{r}}-\frac{\epsilon^2}{2\hat{r}}) \\
&= -\frac{2\epsilon\hat{r}+\epsilon^2}{2\sigma^2}+\frac{2\hat{r}\epsilon-\epsilon^2}{2\sigma^2} \\
&= -\frac{\epsilon^2}{\sigma^2}
\end{aligned}
\tag{19}
$$

3

Therefore,

$$\rho(\hat{r} + \epsilon) = \rho(\hat{r})exp(-\frac{\epsilon^2}{\sigma^2}) \tag{20}$$

Therefore, for $\|x\|^2 = 0$, $\rho(x) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}$, for $\|x\|^2 = \sigma^2 d$, $\rho(x) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}e^{-\frac{d}{2}}$, therefore the probability density is $e^{\frac{d}{2}}$ bigger.

2 Logestic Regression

According to the equation of cost function is

$$E(x) = -\sum_{n=1}^{N}\{t^n ln(y^n) + (1 - t^n)ln(1 - y^n)\} \tag{21}$$

Therefore, the derivative of $E(x)$ is

$$-\frac{\partial E(x)}{\partial w_j} = \sum_{n=1}^{N}\{\frac{t_n}{y^n} \cdot g^{'}(w^T x) \cdot x_j^n - \frac{1 - t^n}{1 - y^n} \cdot g^{'}(w^T x) \cdot x_j^n\}$$

$$= \sum_{n=1}^{N}\{\frac{t^n(1 - y^n) - y^n(1 - t^n)}{y^n(1 - y^n)} g^{'}(w^T x)x_j^n\} \tag{22}$$

$$= \sum_{n=1}^{N}\{\frac{t^n - y^n}{y^n(1 - y^n)} g^{'}(w^T x)x_j^n\}$$

Since logistic activation function is a sigmoid function, one property of the sigmoid function is

$$g^{'}(x) = g(x)(1 - g(x)) \tag{23}$$

Then the denominator of equation (10) can be canceled with $g^{'}(w^T x)$, so the final format of $-\frac{\partial E(x)}{\partial w_j}$ can be transformed to

$$-\frac{E(x)}{w_j} = \sum_{n=1}^{N}\{(t^n - y^n)x_j^n\} \tag{24}$$

4