Prgramming Assignment 1 (Part I)

Huimin Zeng
Department of Computer Science Engineering
University of California, San Diego
San Diego, CA 92122
zenghuimin@yahoo.com

1 Bishop Exercise 1.1

According the problem setting, we have 2 equations available for further computation:

$$\int_{-\infty}^{\infty} e^{-\frac{\lambda}{2}x^2} dx = (\frac{2\pi}{\lambda})^{\frac{1}{2}}$$

$$\Pi_{i=1}^{d} \int_{-\infty}^{\infty} e^{-x_i^2} dx_i = S_d \int_0^{\infty} e^{-r^2} r^{d-1} dr$$

1.1 Deriving S_d

Regarding the first known equation, if we plug in $\lambda = 2$, we obtain

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

Therefore, if we plug in (1) into the second known equation, we have

$$\Pi_{i=1}^{d} \int_{-\infty}^{\infty} e^{-x_i^2} dx_i = (\pi)^{\frac{d}{2}} \tag{2}$$

Moreover, we have a closed-form expression for $\Gamma(x)$. That is

$$\Gamma(x) = \int_0^{\infty} u^{x-1} e^{-u} du$$

If we replace x with $\frac{d}{2}$,

$$\Gamma(\frac{d}{2}) = \int_0^{\infty} u^{\frac{d}{2}} e^{-u} du \tag{3}$$

If we replace $u = r^2$ and therefore $du = 2rdr$ in (3), we have

$$\begin{aligned} \Gamma(\frac{d}{2}) &= \int_0^{\infty} u^{\frac{d}{2}-1} e^{-u} du \\ &= \int_0^{\infty} r^{2(\frac{d}{2}-1)} e^{-r^2} 2rdr \\ &= 2\int_0^{\infty} r^{d-1} e^{-r^2} dr \end{aligned} \tag{4}$$

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

Combining (2) and (4), we complete the proof with following pattern matching:

$$\begin{aligned} S_d \int_0^\infty e^{-r^2} r^{d-1} dr &= S_d \frac{1}{2} (2 \int_0^\infty e^{-r^2} r^{d-1} dr) \\ &= \frac{S_d}{2} \Gamma(\frac{d}{2}) \\ &= \int_{-\infty}^\infty e^{-\frac{\lambda}{2} x^2} dx \\ &= (\pi)^{\frac{d}{2}} \end{aligned} \tag{5}$$

Therefore, we have the result

$$S_d = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \tag{6}$$

1.2 Plugging in, d=2 in (6)

$$S_d = \frac{2\pi^{d/2}}{\Gamma(d/2)} = \frac{2\pi^{2/2}}{\Gamma(2/2)} = \frac{2\pi}{\Gamma(1)} = 2\pi$$

1.3 Plugging in, d=3 in (6)

$$S_d = \frac{2\pi^{d/2}}{\Gamma(d/2)} = \frac{2\pi^{3/2}}{\Gamma(3/2)} = \frac{2\pi^{3/2}}{\sqrt{\pi}/2} = 4\pi$$

2 Bishop Exercise 1.2

2.1 Volume of Hypersphere

A straightforward method to compute the volume of Hypersphere is just to compute the integral.

$$V_d = S_d \int_0^a r^{d-1} dr = \frac{S_d a^d}{d} \tag{7}$$

2.2 Ratio of Volumes

Let's denote the volume of the hypersphere as V_s and the volume of the hypercube as V_c. Therefore, just by plugging the definition, we have

$$\frac{V_s}{V_c} = \frac{\frac{S_d a^d}{d}}{(2a)^d} = \frac{S_d}{d2^d} = \frac{2\pi^{d/2}}{d2^d \Gamma(d/2)} = \frac{\pi^{d/2}}{d2^{d-1}\Gamma(d/2)} \tag{8}$$

2.3 Limit Value

Using the approximation for $\Gamma(x+1)$:

$$\Gamma(x+1) \approx (2\pi)^{\frac{1}{2}} e^{-x} x^{x+\frac{1}{2}}$$

We firstly compute the approximation for $\Gamma(x+1)$:

$$\Gamma(x) \approx (2\pi)^{\frac{1}{2}} e^{-x+1} x^{x-\frac{1}{2}} \tag{9}$$

Similarly, if we replace x with $\frac{d}{2}$ in (9).

$$\Gamma(\frac{d}{2}) \approx (2\pi)^{\frac{1}{2}} e^{-\frac{d}{2}+1} (\frac{d}{2})^{\frac{d}{2}-\frac{1}{2}} \tag{10}$$

Therefore, using the result from (8) and plugging in (10), we have

$$\begin{aligned}
\frac{V_s}{V_c} &= \frac{\pi^{\frac{d}{2}}}{d2^{d-1}\Gamma(\frac{d}{2})} \\
&\approx \frac{\pi^{\frac{d}{2}}}{d2^{d-1}(2\pi)^{\frac{1}{2}} e^{-\frac{d}{2}+1} (\frac{d}{2})^{\frac{d-1}{2}}} \\
&= \frac{(\frac{e\pi}{2})^{\frac{d}{2}}}{C d^{\frac{d}{2}} d^{\frac{1}{2}}} \qquad (C = 2\sqrt{2}e(2\pi)^{\frac{1}{2}}) \\
&= \frac{(\frac{e\pi}{2d})^{\frac{d}{2}}}{C d^{\frac{1}{2}}} \\
&= \frac{(\frac{K}{d})^{\frac{d}{2}}}{C d^{\frac{1}{2}}} \qquad (K = \frac{e\pi}{2})
\end{aligned} \tag{11}$$

For the numerator:

$$\lim_{d\to\infty} log(\frac{K}{d})^{\frac{d}{2}} = \lim_{d\to\infty} \frac{d}{2}(logK - logd) = -\infty$$

$$\Rightarrow \lim_{d\to\infty} (\frac{K}{d})^{\frac{d}{2}} = 0$$

For the denominator:

$$\lim_{d\to\infty} d^{\frac{1}{2}} = \infty$$

In total, we have

$$\lim_{d\to\infty} R(d) = \lim_{d\to\infty} \frac{(\frac{K}{d})^{\frac{d}{2}}}{C d^{\frac{1}{2}}} = \frac{0}{\infty} = 0$$

2.4 Ratio of Distances

Let d_1 denote the distance from the centre of the hypercube to one of the corners, and d_2 is the perpendicular distance to one of the faces. Therefore, by definition, we have

$$\frac{d_1}{d_2} = \frac{(da^2)^{\frac{1}{2}}}{a} = \sqrt{d}$$

3 Bishop Exercise 1.3

3.1 Ratio

Let V_s denote the volume of the hypersphere and V_{shell} denotes the volume of the shell. So we can compute the volume of the shell by simply subtracting the volume of the smaller sphere from the entire sphere.

$$\begin{aligned} V_{shell} &= V_s(a) - V_s(a-\epsilon) \\ &= \frac{S_d a^d}{d} - \frac{S_d (a-\epsilon)^d}{d} \end{aligned} \tag{12}$$

3.2 Limit Value

Hence, we compute the ratio by definition.

$$\frac{V_{shell}}{V_s} = \frac{\frac{S_d a^d}{d} - \frac{S_d (a-\epsilon)^d}{d}}{\frac{S_d a^d}{d}} = 1 - (1 - \frac{\epsilon}{a})^d \tag{13}$$

3.3 Certain Values

Then, we can compute the limit value of the ration as well.

$$\lim_{d\to\infty} 1 - (1 - \frac{\epsilon}{a})^d = 1 \qquad (1 - \frac{\epsilon}{a} \leq 1) \tag{14}$$

With (13), we can simply plug in $\frac{\epsilon}{a} = 0.01$, as well as $d = 2$, $d = 10$, $d = 1000$. The results are demonstrated in the following table.

d	ratio
2	0.0199
10	0.0956
1000	1.0000

With (13), we can simply plug in $\frac{\frac{a}{2}}{a} = 0.5$, as well as $d = 2$, $d = 10$, $d = 1000$. The ration of the volum lies in the $\frac{a}{2}$ results are demonstrated in the following table.

d	ratio
2	0.25
10	0.00098
1000	0

4 Bishop Exercise 1.4

4.1 Cartesian to Polar Coordinates

Based on the knowledge that the probability mass tends to concentrate on the shell with the growth the dimensionality. Therefore, we can compute the integral of the probability dense function on the shell area and approximate it with the integral of the PDF on the entire space. Therefore, starting the equation in the book:

$$\int_{-\infty}^{\infty} p(\boldsymbol{x}) d\boldsymbol{x} = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{||\boldsymbol{x}||^2}{2\sigma^2}}$$

For the sake of simplicity, we denote $\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}$ as Σ^{-1}. And then, we firstly use the approximation and then do some pattern matching.

$$\begin{aligned} \int_{||\boldsymbol{x}||=r-\epsilon}^{||\boldsymbol{x}||=r} p(\boldsymbol{x}) d\boldsymbol{x} &= \int_{-\infty}^{\infty} p(\boldsymbol{x}) d\boldsymbol{x} \\ &= \int_{-\infty}^{\infty} \Sigma^{-1} e^{-\frac{||\boldsymbol{x}||^2}{2\sigma^2}} d\boldsymbol{x} \qquad (\Sigma^{-1} = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}) \end{aligned} \tag{15}$$

If each entry of $\boldsymbol{x}$ is i.i.d., then we are able to do the factorization.

$$\int_{||\boldsymbol{x}||=r-\epsilon}^{||\boldsymbol{x}||=r} p(\boldsymbol{x}) d\boldsymbol{x} = \Sigma^{-1} \Pi_{i=1}^{d} \int_{-\infty}^{\infty} e^{-\frac{x_i^2}{2\sigma^2}} dx_i \tag{16}$$

By replacing $x_i = \sqrt{2}\sigma u_i$ and therefore $dx_i = \sqrt{2}\sigma du_i$, we have further results

$$\int_{||\boldsymbol{x}||=r-\epsilon}^{||\boldsymbol{x}||=r} p(\boldsymbol{x}) d\boldsymbol{x} = (\sqrt{2}\sigma)^d \Sigma^{-1} \Pi_{i=1}^{d} \int_{-\infty}^{\infty} e^{-u_i^2} du_i \qquad (u_i = \frac{x_i}{\sqrt{2}\sigma}) \tag{17}$$

Recall, from the very beginning, we have this equation:

$$\Pi_{i=1}^{d} \int_{-\infty}^{\infty} e^{-x_i^2} dx_i = S_d \int_{0}^{\infty} e^{-r^2} r^{d-1} dr$$

So, we can transform the Cartesian coordinates into polar coordinates.

$$\begin{aligned} \int_{||\boldsymbol{x}||=r-\epsilon}^{||\boldsymbol{x}||=r} p(\boldsymbol{x}) d\boldsymbol{x} &= (\sqrt{2}\sigma)^d \Sigma^{-1} \Pi_{i=1}^{d} \int_{-\infty}^{\infty} e^{-u_i^2} du_i \\ &= (\sqrt{2}\sigma)^d \Sigma^{-1} S_d \int_{0}^{\infty} e^{-v^2} v^{d-1} dv \end{aligned} \tag{18}$$

To derive the final results, we further use $v = \frac{r}{\sqrt{2}\sigma}$ and $dv = d\frac{r}{\sqrt{2}\sigma}$. Then we have

$$\begin{aligned} \int_{||\boldsymbol{x}||=r-\epsilon}^{||\boldsymbol{x}||=r} p(\boldsymbol{x}) d\boldsymbol{x} &= (\sqrt{2}\sigma)^d \Sigma^{-1} S_d \int_{0}^{\infty} e^{-v^2} v^{d-1} dv \\ &= (\sqrt{2}\sigma)^d \Sigma^{-1} S_d \int_{0}^{\infty} e^{-\frac{r^2}{2\sigma^2}} (\frac{r}{\sqrt{2}\sigma})^{d-1} d\frac{r}{\sqrt{2}\sigma} \qquad (v = \frac{r}{\sqrt{2}\sigma}) \end{aligned} \tag{19}$$

With some simple pattern matching, we complete the proof:

$$\begin{aligned} \int_{||\boldsymbol{x}||=r-\epsilon}^{||\boldsymbol{x}||=r} p(\boldsymbol{x}) d\boldsymbol{x} &= \int_{0}^{\infty} \frac{S_d r^{d-1}}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{r^2}{2\sigma^2}} r^{d-1} dr \qquad (\Sigma^{-1} = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}) \\ &= \int_{0}^{\infty} p(r) dr \end{aligned} \tag{20}$$

4.2 Maximum Value

According the setting, we have the expression:

$$p(r) = \frac{S_d r^{d-1}}{(2\pi\sigma^2)^{1/2}} exp(-\frac{r^2}{2\sigma^2})$$

Instead of taking derivative for exponential terms, we can compute all the derivatives in the logarithm space. Since log is a non-decreasing function, it is the operation that preserves the convexity of the original funcion. Therefore, we have

$$\begin{aligned} logp(r) &= log(S_d) + log(r^{d-1}) - \frac{1}{2}log(2\pi\sigma^2) - \frac{r^2}{2\sigma^2} \\ &= log(S_d) + (d-1)log(r) - \frac{1}{2}log(2\pi\sigma^2) - \frac{r^2}{2\sigma^2} \end{aligned} \tag{21}$$

We take the derivative of (21) w.r.t. r,

$$\frac{\partial logp(r)}{\partial r} = \frac{d-1}{r} - \frac{r}{\sigma^2} \tag{22}$$

Set (22) equal to 0, we have

$$\frac{\partial logp(r)}{\partial r} = \frac{d-1}{r} - \frac{r}{\sigma^2} = 0 \tag{23}$$

Therefore,

$$\hat{r} = \sigma\sqrt{d-1} \tag{24}$$

When d grows to infinity, we have

$$\lim_{d\to\infty} \hat{r} = \lim_{d\to\infty} \sigma\sqrt{d-1} = \sigma\sqrt{d} \tag{25}$$

The next step is to verify that r^* is **indeed maximizor**. We manipulate everything in the log-space. That is , we would like to prove the second derivative of the log probability w.r.t. is negative:

$$\frac{\partial^2 logp(r)}{\partial r^2} = \frac{1-d}{r^2} - \frac{1}{\sigma^2} \tag{26}$$

For (26), when $d \to \infty$:

$$\lim_{d\to\infty} \frac{\partial^2 logp(\hat{r})}{\partial \hat{r}^2} = \frac{1-d}{\hat{r}^2} - \frac{1}{\sigma^2} < 0 \tag{27}$$

Therefore, $\hat{r}$ has to be the maximizor.

4.3 Decomposition

The goal:

$$p(\hat{r})e^{-\frac{\epsilon^2}{\sigma^2}} = \frac{S_d\hat{r}^{d-1}}{2\pi\sigma^2} e^{-\frac{\hat{r}^2}{2\sigma^2}} e^{-\frac{\epsilon^2}{\sigma^2}}$$

We start from the beginning and expand it step by step:

$$\begin{aligned} p(\hat{r}+\epsilon) &= \frac{S_d(\hat{r}+\epsilon)^{d-1}}{2\pi\sigma^2} e^{-\frac{(\hat{r}+\epsilon)^2}{2\sigma^2}} \\ &= \frac{S_d\hat{r}^{d-1}(1+\frac{\epsilon}{\hat{r}})^{d-1}}{2\pi\sigma^2} e^{-\frac{\hat{r}^2}{2\sigma^2}} e^{-\frac{\hat{r}\epsilon}{\sigma^2}} e^{-\frac{\epsilon^2}{2\sigma^2}} \\ &= \frac{S_d\hat{r}^{d-1}}{2\pi\sigma^2} e^{-\frac{\hat{r}^2}{2\sigma^2}} e^{-\frac{\epsilon^2}{\sigma^2}} e^{-\frac{\hat{r}\epsilon}{\sigma^2}} e^{log(1+\frac{\epsilon}{\hat{r}})^{d-1}} e^{\frac{\epsilon^2}{2\sigma^2}} \\ &= \frac{S_d\hat{r}^{d-1}}{2\pi\sigma^2} e^{-\frac{\hat{r}^2}{2\sigma^2}} e^{-\frac{\epsilon^2}{\sigma^2}} e^{-\frac{\hat{r}\epsilon}{\sigma^2} + log(1+\frac{\epsilon}{\hat{r}})^{d-1} + \frac{\epsilon^2}{2\sigma^2}} \end{aligned} \tag{28}$$

Compare (28) and the condition, we define a function:

$$f(\epsilon) = -\frac{\hat{r}\epsilon}{\sigma^2} + (d-1)log(1+\frac{\epsilon}{\hat{r}}) + \frac{\epsilon^2}{2\sigma^2} \tag{29}$$

All we need is to prove that (29) is equal to 0. And we complete the proof via Taylor expansion at $\epsilon = 0$:

$$\begin{aligned} f(\epsilon) &\approx f(0) + \frac{\partial f(\epsilon)}{\partial \epsilon}\epsilon + \frac{\partial^2 f(\epsilon)}{\partial \epsilon^2}\frac{\epsilon^2}{2} \\ &= \frac{d-1}{\hat{r}}\epsilon - \frac{\hat{r}}{\sigma^2}\epsilon - \frac{d-1}{2\sigma^2 d}\epsilon^2 + \frac{\epsilon^2}{2\sigma^2} \end{aligned} \tag{30}$$

Plug in $\hat{r} = \sqrt{d}\sigma$, we have

$$\lim_{d\to\infty} = \frac{d}{\sqrt{d}\sigma}\epsilon - \frac{1}{\sqrt{d}\sigma}\epsilon - \frac{\sqrt{d}\sigma}{\sigma^2}\epsilon - \frac{d}{2d\sigma^2}\epsilon^2 + \frac{1}{2d\sigma^2}\epsilon^2 + \frac{\epsilon^2}{2\sigma^2} = 0 \tag{31}$$

Thus, plug in (28):

$$\begin{aligned} p(\hat{r}+\epsilon &= \frac{S_d \hat{r}^{d-1}}{2\pi\sigma^2} e^{-\frac{\hat{r}^2}{2\sigma^2}} e^{-\frac{\epsilon^2}{\sigma^2}} e^{-\frac{\hat{r}\epsilon}{\sigma^2} + log(1+\frac{\epsilon}{\hat{r}})^{d-1} + \frac{\epsilon^2}{2\sigma^2}} \\ &= \frac{S_d \hat{r}^{d-1}}{2\pi\sigma^2} e^{-\frac{\hat{r}^2}{2\sigma^2}} e^{-\frac{\epsilon^2}{\sigma^2}} e^0 \\ &= \frac{S_d \hat{r}^{d-1}}{2\pi\sigma^2} e^{-\frac{\hat{r}^2}{2\sigma^2}} e^{-\frac{\epsilon^2}{\sigma^2}} \\ &= p(\hat{r}) e^{-\frac{\epsilon^2}{\sigma^2}} \end{aligned} \tag{32}$$

5 Exercise 1.5

Given the loss function:

$$E(w) = -\Sigma_{n=1}^{N} \left\{ t^n ln(y^n) + (1-t^n) ln(1-y^n) \right\}$$

$$y = \frac{1}{1+e^{-w^T x}} = g(w^T x)$$

Following the chain rule, we can calculate the derivative of the loss function w.r.t. the parameter w_j:

$$\begin{aligned} -\frac{E(w)}{w_j} &= -\frac{\partial E(w)}{\partial (y^n)} \frac{\partial y^n}{\partial g^n} \frac{\partial g^n}{\partial w_j} \\ &= \Sigma_{n=1}^{N} \left\{ \frac{t^n}{y^n} - \frac{1-t^n}{1-y^n} \right\} 1 y^n (1-y^n) x_j^n \\ &= \Sigma_{n=1}^{N} \left\{ t^n - y^n \right\} x_j^n \end{aligned} \tag{33}$$