PA1 Individual Report

Zhenrui Yue
Computer Science & Engineering
UC San Diego
La Jolla, CA 92093
yuezrhb@gmail.com

1 Problems from Bishop (20 points)

1.1 Problem 1.1

Given the equations involving the transformation from Cartesian to polar coordinates:

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

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

Let $\lambda = 2$, we transform the first equation into:

$$\int_{-\infty}^{\infty} e^{-x^2 dx} = \pi^{1/2}$$

Plug into the left side of the second equation:

$$\pi^{d/2} = S_d \int_0^{\infty} e^{-r^2} r^{d-1} dr$$

By gamma function definition, we substitute u with r^2:

$$\Gamma(d/2) = \int_0^{\infty} u^{\frac{d}{2}-1} e^{-u} du = 2\int_0^{\infty} r^{d-1} e^{-r^2} dr$$

Plug into the right side of the second equation and rearrange the equation, we proved:

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

For $d = 2$:

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

For $d = 3$:

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

1.2 Problem 1.2

Given the surface of a unit hypersphere S_d, the surface of a hypersphere with radius a in d dimensions could be written as:

$$S_d(a) = a^{d-1} S_d$$

The corresponding volume of hypersphere is its integration of surface area from 0 to a:

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

Therefore, the volume of a hypersphere with radius a divided by a hypercube with side length $2a$:

$$\frac{volume\ of\ sphere}{volume\ of\ cube}=\frac{S_d a^d}{d(2a)^d}=\frac{2\pi^{d/2}a^d}{d(2a)^d\Gamma(d/2)}=\frac{\pi^{d/2}}{\mathrm{d}2^{d-1}\Gamma(d/2)}$$

As $d\to\infty$, substitute d with $2x$,

$$\lim_{d\to\infty}\frac{volume\ of\ sphere}{volume\ of\ cube}=\lim_{d\to\infty}\frac{\pi^{\frac{d}{2}}}{\mathrm{d}2^{d-1}\Gamma\left(\frac{d}{2}\right)}$$

$$=\lim_{x\to\infty}\frac{\pi^x}{2x2^{2x-1}\Gamma(x)}$$

$$=\lim_{x\to\infty}\frac{\pi^x}{2x2^{2x-1}(2\pi)^{\frac{1}{2}}e^{-(x-1)}(x-1)^{x-\frac{1}{2}}}$$

$$=\lim_{x\to\infty}\frac{Ke^x\pi^x}{x4^x(x-1)^{x-\frac{1}{2}}},\qquad K\ is\ a\ postive\ constant$$

$$=\lim_{x\to\infty}\frac{Ke^x\pi^x}{x4^x(x-1)^{x-\frac{1}{2}}}$$

The above equation could be seen as the product of two different parts:

$$\lim_{x\to\infty}\frac{K\pi^x}{4^x}=0$$

$$\lim_{x\to\infty}\frac{e^x}{x(x-1)^{x-\frac{1}{2}}}=\lim_{x\to\infty}e^{\log\frac{e^x}{x(x-1)^{x-\frac{1}{2}}}}$$

$$=\lim_{x\to\infty}e^{\left(x-\log x-\left(x-\frac{1}{2}\right)\log(x-1)\right)}$$

$$=e^{\lim\limits_{x\to\infty}\left(x-\log x-\left(x-\frac{1}{2}\right)\log(x-1)\right)}$$

$$=e^{-\infty}=0$$

Therefore,

$$\lim_{d\to\infty}\frac{volume\ of\ sphere}{volume\ of\ cube}=0$$

The distance of hypercube (side length 2a, from center to one of its corners) and hypersphere (radius a, from center to surface) could be written as:

$$distance\ hypercube=\sqrt{\sum_{i=1}^{d}a^2}=\sqrt{da^2}=a\sqrt{d}$$

$$distance\ hypersphere=a$$

Thus, we proved the ratio would be ∞ as $d\to\infty$,

$$\lim_{d\to\infty}\frac{distance\ hypercube}{distance\ hypersphere}=\lim_{d\to\infty}\sqrt{d}=\infty$$

1.3 Problem 1.3

The volume of hypersphere with radius a in d dimension:

$$V_d(a) = \frac{S_d a^d}{d}$$

The volume of hypersphere with radius $a - \varepsilon$ in d dimension:

$$V_d(a-\varepsilon) = \frac{S_d (a-\varepsilon)^d}{d}$$

Hence,

$$f = \frac{volume\ of\ shell}{volume\ of\ sphere} = \frac{V_d(a) - V_d(a-\varepsilon)}{V_d(a)} = \frac{a^d - (a-\varepsilon)^d}{a^d} = 1 - (1 - \frac{\varepsilon}{a})^d$$

For $\frac{\varepsilon}{a} = 0.01$,

$Dimension$	f
$d = 2$	0.0199
$d = 10$	0.0956
$d = 1000$	0.999

$$g = \frac{volume\ of\ sphere\ (\frac{a}{2})}{volume\ of\ sphere} = \frac{V_d\left(\frac{a}{2}\right)}{V_d(a)} = \frac{\left(\frac{a}{2}\right)^d}{a^d} = \left(\frac{1}{2}\right)^d$$

Fraction of the sphere volume which lies inside the radius of $\frac{a}{2}$:

$Dimension$	g
$d = 2$	0.25
$d = 10$	9.766×10^{-4}
$d = 1000$	0

1.4 Problem 1.4

Given the probability density function $p(x)$:

$$p(x) = \frac{1}{(2\pi\sigma)^{1/2}} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right)$$

The probability mass of thin shell with radius r equals the product of sphere surface and probability density with $\|x\|^2 = \|r\|^2$:

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

Hence the probability mass of this thin shell with thickness ε is $p(r)\varepsilon$

The first and second order derivative of $p(r)$ is:

$$p'(r) = \frac{S_d r^{d-2} \exp\left(-\frac{r^2}{2\sigma^2}\right)}{(2\pi\sigma)^{\frac{1}{2}}} \left\{(d-1) - \frac{r^2}{\sigma^2}\right\} = K\left((d-1) - \frac{r^2}{\sigma^2}\right)$$

$$p''(r) = \frac{S_d r^{d-3} \exp\left(-\frac{r^2}{2\sigma^2}\right)}{(2\pi\sigma)^{\frac{1}{2}}} \left\{1 + (d-2)(d-1) - \frac{r^2}{\sigma^2} - \frac{2r^2(d-1)}{\sigma^2}\right\}$$

Let $p'(r) = 0$, for large d we could approximate $d-1$ to d:

$$r = \sigma\sqrt{d-1} \approx \sigma\sqrt{d}$$

Plug this value into the second order derivative:

$$p''\left(\sigma\sqrt{d}\right) = K(-d^2 - 2d + 3)\,, \qquad K\ is\ a\ postive\ constant$$

$p''(\sigma\sqrt{d})$ equals 0 for $d = 1$ and is negative for $d > 1$, therefore we proved the function $p(r)$ has a single maximum for large d at $\hat{r} = \sigma\sqrt{d}$, thus,

$$p(\hat{r}+\varepsilon) = \frac{S_d(\hat{r}+\varepsilon)^{d-1}\, exp\left(-\frac{(\hat{r}+\varepsilon)^2}{2\sigma^2}\right)}{(2\pi\sigma)^{\frac{1}{2}}}$$

$$= \frac{S_d(\hat{r})^{d-1}\left(1+\frac{\varepsilon}{\hat{r}}\right)^{d-1} exp\left(-\frac{\hat{r}^2}{2\sigma^2}-\frac{2\hat{r}\varepsilon+\varepsilon^2}{2\sigma^2}\right)}{(2\pi\sigma)^{\frac{1}{2}}}$$

$$= \frac{S_d(\hat{r})^{d-1}\, exp\left((d-1)\log\left(1+\frac{\varepsilon}{\hat{r}}\right)\right) exp\left(-\frac{\hat{r}^2}{2\sigma^2}\right) exp\left(-\frac{2\hat{r}\varepsilon+\varepsilon^2}{2\sigma^2}\right)}{(2\pi\sigma)^{\frac{1}{2}}}$$

With Taylor expansion of $\log(x)$ at $x = 1$:

$$\log\left(1+\frac{\varepsilon}{\hat{r}}\right) = \log(1) + \log'(1)\left(\frac{\varepsilon}{\hat{r}}\right) + \frac{1}{2}\log''(1)\left(\frac{\varepsilon}{\hat{r}}\right)^2 = \frac{\varepsilon}{\hat{r}} - \frac{\varepsilon^2}{2\hat{r}^2}$$

The above equation could be further simplified:

$$p(\hat{r}+\varepsilon) = \frac{S_d(\hat{r})^{d-1}\, exp\left((d-1)\log\left(1+\frac{\varepsilon}{\hat{r}}\right)\right) exp\left(-\frac{\hat{r}^2}{2\sigma^2}\right) exp\left(-\frac{2r\varepsilon+\varepsilon^2}{2\sigma^2}\right)}{(2\pi\sigma)^{\frac{1}{2}}}$$

$$\approx \frac{S_d(\hat{r})^{d-1}\, exp\left(-\frac{\hat{r}^2}{2\sigma^2}\right) exp\left(d\left(\frac{\varepsilon}{\sigma\sqrt{d}} - \frac{\varepsilon^2}{2\left(\sigma\sqrt{d}\right)^2}\right)\right) exp\left(-\frac{2\varepsilon+\varepsilon^2}{2\sigma^2}\right)}{(2\pi\sigma)^{\frac{1}{2}}}$$

$$\approx p(\hat{r}) exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$$

2 Logistic Regression

Given the Cross-Entropy and parametrized model,

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

$$y(x) = \frac{1}{1+\exp\left(-\omega^T x\right)}$$

Note $y(x)$ is the sigmoid function and thus:

$$y'(x) = y(x)(1-y(x))$$

The negated partial derivative could be simplified correspondingly:

$$-\frac{\partial E(\omega)}{\partial \omega_j} = \sum_{n=1}^{N}\left\{\frac{t^n}{y^n} y^n(1-y^n)x_j^n - \frac{1-t^n}{1-y^n} y^n(1-y^n)x_j^n\right\}$$

$$= \sum_{n=1}^{N}\{t^n(1-y^n)x_j^n - (1-t^n)y^n x_j^n\}$$

$$= \sum_{n=1}^{N}\{t^n - y^n\}\, x_j^n$$

Hence, we proved the partial gradient of the Cross-Entropy.