Math 789 Assignment 2

Brendan Drachler

January 30, 2020

Bishop 1.1

Starting from,

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

Our goal is to solve for S_d and make sure it agrees with the analytical expression in Eq. 1.43. When d=2, the expression becomes (thanks Wolfram!):

$$\prod_{i=1}^{2} \int_{-\infty}^{\infty} e^{x_i^2} dx_i = S_2 \int_0^{\infty} e^{-r^2} r dr = \frac{1}{2} S_2 \tag{2}$$

$$\downarrow \tag{3}$$

$$(\int_{-\infty}^{\infty} e^{x_1^2} dx_1)(\int_{-\infty}^{\infty} e^{x_2^2} dx_2) = \frac{1}{2} S_2 \tag{4}$$

Each of the integrals on the left hand side evaluates to $\sqrt{\pi}$ (thanks Wolfram!). Therefore, we are left with:

$$2\pi = S_2 \tag{5}$$

Now, this should be reproducible by plugging $d = 2$ into Eq. 1.43.

$$S_2 = \frac{2\pi}{\Gamma(\frac{2}{2})} = 2\pi \tag{6}$$

Validated! Moving onto d=3,

$$(\int_{-\infty}^{\infty} e^{x_1^2} dx_1)(\int_{-\infty}^{\infty} e^{x_2^2} dx_2)(\int_{-\infty}^{\infty} e^{x_3^2} dx_3) = S_3 \int_0^{\infty} e^{-r^2} r^2 dr = \frac{\sqrt{\pi}}{4} S_3 \tag{7}$$

$$\downarrow \tag{8}$$

$$\pi^{\frac{3}{2}} = \frac{\pi^{\frac{1}{2}}}{4} S_3 \rightarrow S_3 = 4\pi \tag{9}$$

Once again, this should be reproducible by plugging d=3 into Eq. 1.43.

$$S_3 = \frac{2\pi^{\frac{3}{2}}}{\Gamma(\frac{3}{2})} = \frac{2\pi^{\frac{3}{2}}}{\frac{\pi^{\frac{1}{2}}}{2}} = 4\pi \tag{10}$$

Bishop 1.3

First, I will define three quantities:

$$V_d = \frac{S_d a^d}{d} \tag{11}$$

$$V_d|_{r=a-\epsilon} = \frac{S_d (a-\epsilon)^d}{d} \tag{12}$$

$$V_d|_{r=\frac{a}{2}} = \frac{S_d (\frac{a}{2})^d}{d} \tag{13}$$

Now, our goal is to find the ratio of the volume of the sphere at $r = a$ to the volume at $r = a - \epsilon$.

$$\frac{V_d|_{r=a-\epsilon}}{V_d} = 1 - (1 - \frac{\epsilon}{a})^d = f \tag{14}$$

We can evaluate f, with $\frac{\epsilon}{a} = 0.01$, for $d = 2, d = 10, d = 1000$:

$$f_{d=2} \approx 0.0199, f_{d=10} \approx 0.0956, f_{d=1000} \approx 0.99996 \tag{15}$$

Now, we can calculate:

$$\frac{V_d|_{r=\frac{a}{2}}}{V_d} = 2^{-d} = f \tag{16}$$

We can evaluate f for $d = 2, d = 10, d = 1000$:

$$f_{d=2} \approx 0.25, f_{d=10} \approx 0.000977, f_{d=1000} \approx 0 \tag{17}$$

Bishop 1.4

Starting from:

$$p(x) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{(-\frac{||x||^2}{2\sigma^2})} \tag{18}$$

By using Eq. 1.42, we can transform this to polar coordinates and yield:

$$p(r) = \frac{S_d r^{d-1}}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{(-\frac{r^2}{2\sigma^2})} \tag{19}$$

Set the derivative equal to 0 and solve for r to find the maximum:

$$0 = \frac{S_d}{2\pi\sigma^2}[(d-1)r^{d-2}e^{-\frac{r^2}{2\sigma^2}} - \frac{r^d}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}}] \tag{20}$$

$$\downarrow \tag{21}$$

$$0 = (d-1)r^{d-2} - \frac{r^d}{\sigma^2} \tag{22}$$

$$\downarrow \tag{23}$$

$$r = \sqrt{(d-1)\sigma^2} \tag{24}$$

If $d >> 1$, this simplifies to $\approx \sqrt{d\sigma^2}$. We can now calculate the ratio of $p(r+\epsilon)$ to $p(r)$ as requested.

$$\frac{p(r+\epsilon)}{p(r)} = \frac{(r+\epsilon)^{d-1}e^{-\frac{(r+\epsilon)^2}{2\sigma^2}}}{(r)^{d-1}e^{-\frac{(r)^2}{2\sigma^2}}} \tag{25}$$

$$\downarrow \tag{26}$$

$$(1+\frac{\epsilon}{r})^{d-1}e^{-\frac{2\epsilon r+\epsilon^2}{2\sigma^2}} \tag{27}$$

$$\downarrow \tag{28}$$

$$e^{(-\frac{2\epsilon r+\epsilon^2}{2\sigma^2}+(d-1)\ln(1+\frac{\epsilon}{r}))} \tag{29}$$

Taylor expanding this around $r = 0$ results in (thanks Wolfram!):

$$p(r+\epsilon) = r^{d-1}e^{-\frac{-r^2}{2\sigma^2}}e^{-\frac{3\epsilon^2}{2\sigma^2}} = p(r)e^{-\frac{3\epsilon^2}{2\sigma^2}} \tag{30}$$

1 Bishop 1.5

We will start by differentiating Bishop Eq. 1.3:

$$E = \frac{1}{2}\sum_{n=1}^{N}[y(x_n;w) - t_n]^2 \tag{31}$$

where $y(x_n;w)$ is given by:

$$\sum_{m=1}^{M} = w_m x^m \tag{32}$$

We want to minimize the weighs, w:

$$\frac{\partial E}{\partial w} = 0 = \sum_{n=1}^{N} [(\sum_{m=0}^{M} w_m x_n^m) x_n^i - x_n^i t_n] \tag{33}$$

$$\downarrow \tag{34}$$

$$\sum_{n=1}^{N} \sum_{m=0}^{M} x_n^{m+i} w_m = \sum_{n=1}^{N} x_n^i t_n \tag{35}$$

Writing $A_{jj'} = \sum_n (x^n)^{j+j'}$ and $T_{j'} = \sum_n t^n (x^n)^{j'}$ as defined in Bishop Eq. 1.53 it is easy to see that the last equation simplifies to Bishop Eq. 1.52:

$$\sum_{m=1}^{M} A_{jj'} w_j = T_{j'} \tag{36}$$

Bishop 1.9

Using Bayes Theorem,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{37}$$

$$\downarrow \tag{38}$$

$$P(Box_1|Apple) = \frac{P(Apple|Box_1)P(Box_1)}{P(Apple)} \tag{39}$$

$$P(Box_1|Apple) = \frac{\frac{8}{12}\frac{1}{2}}{\frac{1}{2}\frac{8}{12} + \frac{10}{12}\frac{1}{2}} = 0.4\bar{4} \tag{40}$$

2 Bishop 1.10

The first bit is trivial (maybe too trivial...)

$$a \leq b \tag{41}$$

Multiply both sides by a and take a square root...

$$a \leq (ab)^{(\frac{1}{2})} \tag{42}$$

The second part isn't as simple:

$$p(error) = \int_{r_1} p(x, C_2)dx + \int_{r_2} p(x, C_1)dx \tag{43}$$

$$\downarrow \tag{44}$$

Using our previous result, we can say:

$$\int_{r_1} p(x, C_2)dx \leq \int_{r_1} [p(x, C_1)p(x, C_2)]^{\frac{1}{2}} dx \tag{45}$$

$$\int_{r_2} p(x, C_1)dx \leq \int_{r_2} [p(x, C_2)p(x, C_1)]^{\frac{1}{2}} dx \tag{46}$$

Combining these results in the same fashion as our trivial example leads to:

$$P(error) \leq [p(x|C_1)P(C_1)p(x|C_2)P(C_2)]^{\frac{1}{2}} \tag{47}$$

Bishop 1.11

I'm not sure I fully understand the question. However, if

$$L_{kj} = 1 - \delta_{kj}, \tag{48}$$

there will be no loss. So by definition it will minimize the probability of misclassification.