

# My Solutions for Exercises of Deep Learning Fundamentals by Bishop

Andres Espinosa

May 11, 2024

## Contents

<b>1</b>	<b>The Deep Learning Revolution</b>	<b>3</b>
1.1	No Exercises . . . . .	3
<b>2</b>	<b>Probabilities</b>	<b>4</b>
2.1	Exercise 2.2 . . . . .	4
2.2	Exercise 2.4 . . . . .	4
2.3	Exercise 2.6 . . . . .	5
2.4	Exercise 2.7 . . . . .	5
2.5	Exercise 2.8 . . . . .	5
2.6	Exercise 2.9 . . . . .	5
2.7	Exercise 2.10 . . . . .	5
2.8	Exercise 2.11 . . . . .	6
2.9	Exercise 2.12 . . . . .	6
2.10	Exercise 2.13 . . . . .	6
2.11	Exercise 2.14 . . . . .	6
2.12	Exercise 2.15 . . . . .	7
2.13	Exercise 2.16 . . . . .	7
2.14	Exercise 2.17 . . . . .	7
2.15	Exercise 2.18 . . . . .	8
2.16	Exercise 2.19 . . . . .	8
2.17	Exercise 2.20 . . . . .	8
2.18	Exercise 2.21 . . . . .	9
2.19	Exercise 2.22 . . . . .	9
2.20	Exercise 2.23 . . . . .	9
2.21	Exercise 2.24 . . . . .	9
2.22	Exercise 2.25 . . . . .	9
2.23	Exercise 2.26 . . . . .	9
2.24	Exercise 2.27 . . . . .	9
2.25	Exercise 2.28 . . . . .	9
2.26	Exercise 2.29 . . . . .	9
2.27	Exercise 2.30 . . . . .	10
2.28	Exercise 2.31 . . . . .	10
2.29	Exercise 2.32 . . . . .	10
2.30	Exercise 2.33 . . . . .	10

2.31	Exercise 2.34	10
2.32	Exercise 2.35	10
2.33	Exercise 2.36	10
2.34	Exercise 2.37	10
2.35	Exercise 2.38	10
2.36	Exercise 2.39	10
2.37	Exercise 2.40	10
2.38	Exercise 2.41	10
<b>3</b>	<b>Standard Distributions</b>	<b>11</b>
3.1	Exercise 3.1	11
3.2	Exercise 3.2	11
3.3	Exercise 3.3	12
3.4	Exercise 3.4	13
3.5	Exercise 3.5	13
3.6	Exercise 3.6	13
3.7	Exercise 3.7	13
3.8	Exercise 3.8	14
3.9	Exercise 3.9	14
3.10	Exercise 3.10	14
3.11	Exercise 3.11	14
3.12	Exercise 3.12	14
3.13	Exercise 3.13	14
3.14	Exercise 3.14	15
3.15	Exercise 3.15	15
3.16	Exercise 3.16	15
3.17	Exercise 3.17	15
3.18	Exercise 3.18	15
3.19	Exercise 3.19	15
3.20	Exercise 3.20	15
3.21	Exercise 3.21	15
3.22	Exercise 3.22	15
3.23	Exercise 3.23	15
3.24	Exercise 3.24	15
3.25	Exercise 3.25	15
3.26	Exercise 3.26	15
3.27	Exercise 3.27	16
3.28	Exercise 3.28	16
3.29	Exercise 3.29	16
3.30	Exercise 3.30	16
3.31	Exercise 3.31	16
3.32	Exercise 3.32	16
3.33	Exercise 3.33	16
3.34	Exercise 3.34	16
3.35	Exercise 3.35	16
3.36	Exercise 3.36	16
3.37	Exercise 3.37	16
3.38	Exercise 3.38	17

# 1 The Deep Learning Revolution

## 1.1 No Exercises

## 2 Probabilities

### Exercise 2.1

Bayes rule

$$P[C = 1|T = 1] = \frac{P[T = 1|C = 1] * P[C = 1]}{P[T = 1|C = 1] * P[C = 1] + P[T = 1|C = 0] * P[C = 0]} \quad (1)$$

$$P[C = 1|T = 1] = \frac{0.90 * 0.001}{0.90 * 0.001 + 0.03 * 0.999} = 0.0292 \quad (2)$$

Given that the test result was positive, there is a 2.92% chance that you have cancer.

### 2.1 Exercise 2.2

Not attempted

### Exercise 2.3

$$p(\mathbf{y}) = \int p_{\mathbf{u}, \mathbf{v}}(\mathbf{u}, \mathbf{y} - \mathbf{u}) d\mathbf{u} \quad (3)$$

$$= \int p_{\mathbf{u}}(\mathbf{u}) p_{\mathbf{v}}(\mathbf{y} - \mathbf{u}) d\mathbf{u} \quad (4)$$

### 2.2 Exercise 2.4

Not attempted

### Exercise 2.5

Exponential:

$$p(x|\lambda) = \lambda e^{-\lambda x} \quad (5)$$

Laplace:

$$p(x|\mu, \gamma) = \frac{1}{2\gamma} e^{-\frac{|x-\mu|}{\gamma}} \quad (6)$$

Verifying that the exponential distribution is normalized:

$$p(x|\lambda) = \lambda e^{-\lambda x} \quad (7)$$

$$\int_0^{\infty} \lambda e^{-\lambda x} = -e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{e^{\infty}} + \frac{1}{e^0} \quad (8)$$

$$= 1 \quad (9)$$

Verifying the laplace distribution:

$$p(x|\mu, \gamma) = \frac{1}{2\gamma} e^{-\frac{|x-\mu|}{\gamma}} \quad (10)$$

$$\begin{cases} \frac{1}{2\gamma} e^{-\frac{x-\mu}{\gamma}} & \text{if } x \geq \mu \\ \frac{1}{2\gamma} e^{-\frac{-x+\mu}{\gamma}} & \text{if } x < \mu \end{cases} \quad (11)$$

$$\int_{\mu}^{\infty} \frac{1}{2\gamma} e^{-\frac{x-\mu}{\gamma}} = \frac{1}{2} e^{-\frac{x-\mu}{\gamma}} \Big|_{\mu}^{\infty} = -\frac{1}{2} (e^{-\infty} - e^0) \quad (12)$$

$$= \frac{1}{2} \quad (13)$$

$$\int_{-\infty}^{\mu} \frac{1}{2\gamma} e^{-\frac{-x+\mu}{\gamma}} = \frac{1}{2} e^{-\frac{-x+\mu}{\gamma}} \Big|_{-\infty}^{\mu} = \frac{1}{2} (e^0 - e^{-\infty}) \quad (14)$$

$$= \frac{1}{2} \quad (15)$$

$$\frac{1}{2} + \frac{1}{2} = 1 \quad (16)$$

### 2.3 Exercise 2.6

Not attempted

### 2.4 Exercise 2.7

$$P(x|D) = \frac{1}{N} \sum_{n=1}^N \delta(x - x_n) \quad (17)$$

$$E[f] = \int p(x) f(x) dx \quad (18)$$

$$\text{Substituting:} \quad (19)$$

$$E[f] = \int \frac{1}{N} \sum_{n=1}^N \delta(x - x_n) f(x) dx \quad (20)$$

$$E[f] = \frac{1}{N} \sum_{n=1}^N \int_{x_n-\varepsilon}^{x_n+\varepsilon} \delta(x - x_n) f(x) dx \quad (21)$$

$$E[f] = \frac{1}{N} \sum_{n=1}^N f(x_n) \int_{x_n-\varepsilon}^{x_n+\varepsilon} \delta(x - x_n) dx \quad (22)$$

$$E[f] = \frac{1}{N} \sum_{n=1}^N f(x_n) \quad (23)$$

### 2.5 Exercise 2.8

Not attempted

### 2.6 Exercise 2.9

$$\text{cov}[X, y] = E_{x,y}[xy] - E[x]E[y] \quad (24)$$

If  $x$  and  $y$  are independent, the joint distribution is equal to the product of the marginals.  $p(x, y) = p(x)p(y)$ . If  $E_{x,y}[xy] = E[x]E[y]$ , then the covariance will be zero.

### 2.7 Exercise 2.10

Not attempted

## 2.8 Exercise 2.11

Proving  $E[x] = E_y[E_x[x|y]]$ :

$$E_x[x|y] = \int p(x|y)xdx \quad (25)$$

$$\text{Substituting:} \quad (26)$$

$$E[x] = E_y[\int p(x|y)xdx] \quad (27)$$

$$E[x] = \int E_y[p(x|y)]xdx \quad (28)$$

$$E[x] = \int \int p(x|y)xp(y)dxdy \quad (29)$$

$$E[x] = \int \int \frac{p(x,y)}{p(y)}xp(y)dxdy \quad (30)$$

$$E[x] = \int \int p(x,y)dxdy \quad (31)$$

$$E[x] = E[x] \quad (32)$$

## 2.9 Exercise 2.12

Not attempted

## 2.10 Exercise 2.13

$$E[x] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma^2} e^{\frac{1}{2\sigma^2}(x-\mu)^2} dx \quad (33)$$

$$\text{Change of variables } z = \frac{x-\mu}{\sigma}, \sigma dz = dx \quad (34)$$

$$E[x] = \int_{-\infty}^{\infty} \sigma \frac{\sigma z + \mu}{\sqrt{2\pi}\sigma^2} e^{\frac{1}{2}z^2} dz \quad (35)$$

$$E[x] = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ze^{\frac{1}{2}z^2} + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{1}{2}z^2} \quad (36)$$

$$E[x] = \frac{\sigma}{\sqrt{2\pi}} * 0 + \frac{\mu}{\sqrt{2\pi}} * \sqrt{2\pi} \quad (37)$$

$$E[x] = \mu \quad (38)$$

## 2.11 Exercise 2.14

Not attempted

### 2.12 Exercise 2.15

Solving for  $\mu_{ml}$ :

$$\log p(x|\mu, \sigma^2) = \frac{-1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2 - \frac{N}{2} \log \sigma^2 - \frac{N}{2} \log 2\pi \quad (39)$$

$$\frac{d}{d\mu} \log p(x|\mu, \sigma^2) = \frac{1}{2\sigma^2} \sum_{n=1}^N 2(x_n - \mu) \quad (40)$$

$$0 = \frac{1}{\sigma^2} \sum_{n=1}^N (x_n - \mu) \quad (41)$$

$$0 = \sum_{n=1}^N x_n - \sum_{n=1}^N \mu \quad (42)$$

$$N\mu = \sum_{n=1}^N x_n \quad (43)$$

$$\mu_{ml} = \frac{1}{N} \sum_{n=1}^N x_n \quad (44)$$

Solving for  $\sigma_{ml}$ :

$$\log p(x|\mu, \sigma^2) = \frac{-1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2 - \frac{N}{2} \log \sigma^2 - \frac{N}{2} \log 2\pi \quad (45)$$

$$\frac{d}{d\sigma^2} \log p(x|\mu, \sigma^2) = \frac{1}{2\sigma^4} \sum_{n=1}^N (x_n - \mu)^2 - \frac{N}{2\sigma^2} \quad (46)$$

$$\frac{N}{2\sigma^2} = \frac{1}{2\sigma^4} \sum_{n=1}^N (x_n - \mu)^2 \quad (47)$$

$$\sigma_{ml}^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_{ml})^2 \quad (48)$$

### 2.13 Exercise 2.16

not attempted

### 2.14 Exercise 2.17

Finding expectation of  $\hat{\sigma}^2$

$$E[\hat{\sigma}^2] = E\left[\frac{1}{N} \sum_{n=1}^N (x_n - \mu)^2\right] \quad (49)$$

$$= \frac{1}{N} \sum_{n=1}^N E[x_n^2 - 2x_n\mu + \mu^2] \quad (50)$$

$$= \frac{1}{N} \sum_{n=1}^N E[x_n^2] - E[2x_n\mu] + E[\mu^2] \quad (51)$$

$$= \frac{1}{N} \sum_{n=1}^N E[x_n^2] - 2E[x_n]E[\mu] + E[\mu^2] \quad (52)$$

$$= \frac{1}{N} \sum_{n=1}^N E[x_n^2] - E[x_n]^2 \quad (53)$$

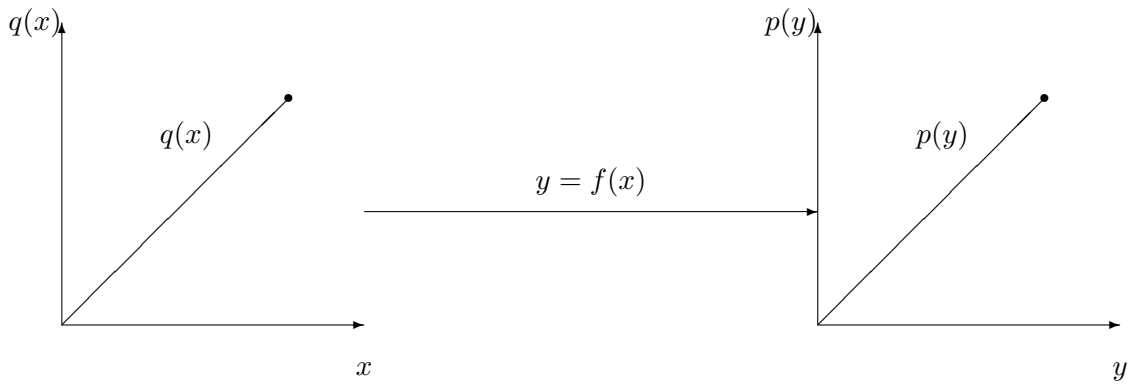
$$= \frac{1}{N} \sum_{n=1}^N \mu^2 + \sigma^2 - \mu^2 \quad (54)$$

$$= \sigma^2 \quad (55)$$

### 2.15 Exercise 2.18

Not attempted

### 2.16 Exercise 2.19



### 2.17 Exercise 2.20

Not attempted



### 2.18 Exercise 2.21

Showing  $h(p^2) = 2h(p)$ :

$$h(p) = h(p(x_1)) + h(p(x_2)) + \cdots + h(p(x_n)) \quad (56)$$

$$h(p^2) = h(x_1^2) + h(x_2^2) + \cdots + h(x_n^2) \quad (57)$$

$$\because h(x) = -\log_2 p(x), \quad (58)$$

$$h(p^2) = 2h(x_1) + 2h(x_2) + \cdots + 2h(x_n) \quad (59)$$

$$h(p^2) = 2h(p) \quad (60)$$

This can be applied to any exponent which includes any choice of  $n$  integer or  $\frac{n}{m}$  positive rational number.

$$h(p^x) = xh(p) \quad \forall Q^+ \quad (61)$$

$$\therefore \quad (62)$$

$$h(p) \propto \ln p \quad (63)$$

### 2.19 Exercise 2.22

Not attempted

### 2.20 Exercise 2.23

Not attempted

### 2.21 Exercise 2.24

Not attempted

### 2.22 Exercise 2.25

Not attempted

### 2.23 Exercise 2.26

Not attempted

### 2.24 Exercise 2.27

Not attempted

### 2.25 Exercise 2.28

Not attempted

### 2.26 Exercise 2.29

Not attempted

**2.27 Exercise 2.30**

Not attempted

**2.28 Exercise 2.31**

Not attempted

**2.29 Exercise 2.32**

Not attempted

**2.30 Exercise 2.33**

Not attempted

**2.31 Exercise 2.34**

Not attempted

**2.32 Exercise 2.35**

Not attempted

**2.33 Exercise 2.36**

Not attempted

**2.34 Exercise 2.37**

Not attempted

**2.35 Exercise 2.38**

Not attempted

**2.36 Exercise 2.39**

Not attempted

**2.37 Exercise 2.40**

Not attempted

**2.38 Exercise 2.41**

Not attempted

### 3 Standard Distributions

#### 3.1 Exercise 3.1

Proving  $\sum_{x=0}^1 p(x|\mu) = 1$ :

$$\text{Bern}(x|\mu) = \mu^x(1-\mu)^{1-x} \quad (64)$$

$$\sum_{x=0}^1 p(x|\mu) = \mu^0(1-\mu)^1 + \mu^1(1-\mu)^0 \quad (65)$$

$$= 1 - \mu + \mu \quad (66)$$

$$= 1 \quad (67)$$

Proving  $E[x] = \mu$ :

$$E[x] = \sum_{x=0}^1 p(x|\mu) * x \quad (68)$$

$$E[x] = 0 * \mu^0(1-\mu)^1 + 1 * \mu^1(1-\mu)^0 \quad (69)$$

$$= 0 * (1-\mu) + 1 * \mu \quad (70)$$

$$= \mu \quad (71)$$

Proving  $\text{var}[x] = \mu(1-\mu)$ :

$$\text{Var}[x] = \sum_i (x_i - \mu)^2 p(x_i) \quad (72)$$

$$= (0 - \mu)^2(1-\mu) + (1 - \mu)^2\mu \quad (73)$$

$$= \mu^2(1-\mu) + (1-\mu)^2\mu \quad (74)$$

$$= \mu(1-\mu)(\mu + (1-\mu)) \quad (75)$$

$$= \mu(1-\mu)(1) \quad (76)$$

$$= \mu(1-\mu) \quad (77)$$

Solving for entropy of the bernoulli distribution proves simple as there are two probabilities,  $p_1 = \mu, p_2 = 1 - \mu$ . Therefore the entropy must be:

$$H[x] = -\mu \ln \mu - (1-\mu) \ln (1-\mu) \quad (78)$$

#### 3.2 Exercise 3.2

Not attempted

### 3.3 Exercise 3.3

Normalizing the binomial distribution  $\binom{N}{m}\mu^m(1-\mu)^{N-m}$ :

$$\text{First step is proving } \binom{N}{m} + \binom{N}{m-1} = \binom{N+1}{m} : \quad (79)$$

$$\binom{N}{m} + \binom{N}{m-1} = \frac{N!}{(N-m)!m!} + \frac{N!}{(N-m+1)!(m-1)!} \quad (80)$$

$$= \frac{N!}{(N-m)!m!} * \frac{N-m+1}{N-m+1} + \frac{N!}{(N-m+1)!(m-1)!} * \frac{m}{m} \quad (81)$$

$$= \frac{N!(N+1-m)}{(N+1-m)!m!} + \frac{N!m}{(N-m+1)!m!} \quad (82)$$

$$= \frac{(N+1)! - N!m + N!m}{(N+1-m)!m!} \quad (83)$$

$$= \frac{(N+1)!}{(N+1-m)!m!} \quad (84)$$

$$= \binom{N+1}{m} \quad (85)$$

With this information we would like to prove that  $(1+x)^N = \sum_{m=0}^N \binom{N}{m}x^m$   
N=2:

$$\binom{2}{0}x^0 + \binom{2}{1}x^1 + \binom{2}{2}x^2 = 1 + 2x + x^2 = (1+x)^2 \quad (86)$$

N=3:

$$\binom{3}{0}x^0 + \binom{3}{1}x^1 + \binom{3}{2}x^2 + \binom{3}{3}x^3 = 1 + 3x + 3x^2 + x^3 = (1+x)^3 \quad (87)$$

N=4

$$\binom{4}{0}x^0 + \binom{4}{1}x^1 + \binom{4}{2}x^2 + \binom{4}{3}x^3 + \binom{4}{4}x^4 = 1 + 4x + 6x^2 + 4x^3 + x^4 = (1+x)^4 \quad (88)$$

By induction:  $\forall x \in \mathbb{R}$  and  $N \in \mathbb{N}$

$$(1+x)^N = \sum_{m=0}^N \binom{N}{m}x^m \quad (89)$$

Normalizing the binomial distribution:

$$\sum_{m=0}^N \binom{N}{m} \mu^m (1-\mu)^{N-m} = 1 \quad (90)$$

$$= (1-\mu)^N \sum_{m=0}^N \binom{N}{m} \mu^m (1-\mu)^{-m} \quad (91)$$

$$= (1-\mu)^N \sum_{m=0}^N \binom{N}{m} \left(\frac{\mu}{1-\mu}\right)^m \quad (92)$$

$$= (1-\mu)^N \left(1 + \frac{\mu}{1-\mu}\right)^N \quad (93)$$

$$= ((1-\mu)(1 + \frac{\mu}{1-\mu}))^N \quad (94)$$

$$= (1 + \frac{\mu}{1-\mu} - \mu - \frac{\mu^2}{1-\mu}) \quad (95)$$

$$= ((1-\mu) + \frac{\mu}{1-\mu}(1-\mu))^N \quad (96)$$

$$= (1-\mu + \mu)^N \quad (97)$$

$$= 1 \quad (98)$$

### 3.4 Exercise 3.4

Not attempted

### 3.5 Exercise 3.5

Finding the mode of the multivariate gaussian. The mode of a distribution is the same as the maximum of the probability density function. For the multivariate gaussian, this is:

$$\mathcal{N}(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)} \quad (99)$$

$$\frac{d}{d\mathbf{x}} \mathcal{N}(\mathbf{x}|\mu, \Sigma) \text{ ignoring normalization constant for brevity} \quad (100)$$

$$0 = -\frac{1}{2} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)} * \frac{d}{d\mathbf{x}} (\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu) \quad (101)$$

$$0 = -\frac{1}{2} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)} * 2\Sigma^{-1}(\mathbf{x}-\mu) \quad (102)$$

$$\mathbf{x} = \mu \quad (103)$$

### 3.6 Exercise 3.6

Not attempted

### 3.7 Exercise 3.7

Finding Kullback-Leibler divergence between  $q(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu_{\mathbf{q}}, \Sigma_{\mathbf{q}})$  and  $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu_{\mathbf{p}}, \Sigma_{\mathbf{p}})$

$$KL(p||q) = - \int p(x) \ln \frac{q(x)}{p(x)} dx \quad (104)$$

$$KL(q(x)||p(x)) = - \int \mathcal{N}(\mathbf{x}|\mu_{\mathbf{q}}, \Sigma_{\mathbf{q}}) \ln \frac{\mathcal{N}(\mathbf{x}|\mu_{\mathbf{p}}, \Sigma_{\mathbf{p}})}{\mathcal{N}(\mathbf{x}|\mu_{\mathbf{q}}, \Sigma_{\mathbf{q}})} \quad (105)$$

$$= - \int \mathcal{N}(\mathbf{x}|\mu_{\mathbf{q}}, \Sigma_{\mathbf{q}}) (\ln \mathcal{N}(\mathbf{x}|\mu_{\mathbf{p}}, \Sigma_{\mathbf{p}}) - \ln \mathcal{N}(\mathbf{x}|\mu_{\mathbf{q}}, \Sigma_{\mathbf{q}})) \quad (106)$$

$$\begin{aligned} \text{definition: } \ln \mathcal{N}(\mathbf{x}|\mu_{\mathbf{p}}, \Sigma_{\mathbf{p}}) &= -\frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_{\mathbf{p}}| - \frac{1}{2} (\mathbf{x} - \mu_{\mathbf{p}})^T \Sigma_{\mathbf{p}}^{-1} (\mathbf{x} - \mu_{\mathbf{p}}) \\ &= -\frac{1}{2} \int \mathcal{N}(\mathbf{x}|\mu_{\mathbf{q}}, \Sigma_{\mathbf{q}}) (-\ln |\Sigma_{\mathbf{p}}| + \ln |\Sigma_{\mathbf{q}}| - (\mathbf{x} - \mu_{\mathbf{p}})^T \Sigma_{\mathbf{p}}^{-1} (\mathbf{x} - \mu_{\mathbf{p}}) + (\mathbf{x} - \mu_{\mathbf{q}})^T \Sigma_{\mathbf{q}}^{-1} (\mathbf{x} - \mu_{\mathbf{q}})) \end{aligned} \quad (107)$$

$$\text{Splitting integral into three} \quad (109)$$

$$\text{First: } -\frac{1}{2} \ln \frac{|\Sigma_{\mathbf{q}}|}{|\Sigma_{\mathbf{p}}|} \int \frac{1}{(2\pi)^{D/2} |\Sigma_{\mathbf{q}}|^{\frac{1}{2}}} e^{-\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{q}})^T \Sigma_{\mathbf{q}}^{-1} (\mathbf{x} - \mu_{\mathbf{q}})} \quad (110)$$

$$\text{Second: } \frac{1}{2} \int \frac{1}{(2\pi)^{D/2} |\Sigma_{\mathbf{q}}|^{\frac{1}{2}}} e^{-\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{q}})^T \Sigma_{\mathbf{q}}^{-1} (\mathbf{x} - \mu_{\mathbf{q}})} (\mathbf{x} - \mu_{\mathbf{p}})^T \Sigma_{\mathbf{p}}^{-1} (\mathbf{x} - \mu_{\mathbf{p}}) \quad (111)$$

$$\text{Third: } -\frac{1}{2} \int \frac{1}{(2\pi)^{D/2} |\Sigma_{\mathbf{q}}|^{\frac{1}{2}}} e^{-\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{q}})^T \Sigma_{\mathbf{q}}^{-1} (\mathbf{x} - \mu_{\mathbf{q}})} (\mathbf{x} - \mu_{\mathbf{p}})^T \Sigma_{\mathbf{p}}^{-1} (\mathbf{x} - \mu_{\mathbf{p}}) \quad (112)$$

$$\text{Evaluating first integral: } -\frac{1}{2} \ln \frac{|\Sigma_{\mathbf{q}}|}{|\Sigma_{\mathbf{p}}|} * 1 \quad (113)$$

$$\text{Evaluating second integral: } \frac{1}{2} (\mu_{\mathbf{p}} - \mu_{\mathbf{q}})^T \Sigma_{\mathbf{p}}^{-1} (\mu_{\mathbf{p}} - \mu_{\mathbf{q}}) \quad (114)$$

$$\text{Evaluating third integral: } -\frac{1}{2} (\mu_{\mathbf{p}} - \mu_{\mathbf{q}})^T \Sigma_{\mathbf{p}}^{-1} (\mu_{\mathbf{p}} - \mu_{\mathbf{q}}) + \frac{1}{2} \text{Tr}(\Sigma_{\mathbf{p}}^{-1} \Sigma_{\mathbf{q}}) \quad (115)$$

I am not sure about this one it is quite hard

### 3.8 Exercise 3.8

Not attempted

### 3.9 Exercise 3.9

Not attempted

### 3.10 Exercise 3.10

Not attempted

### 3.11 Exercise 3.11

Not attempted

### 3.12 Exercise 3.12

Not attempted

### 3.13 Exercise 3.13

Not attempted

**3.14 Exercise 3.14**

Not attempted

**3.15 Exercise 3.15**

Not attempted

**3.16 Exercise 3.16**

Not attempted

**3.17 Exercise 3.17**

Not attempted

**3.18 Exercise 3.18**

Not attempted

**3.19 Exercise 3.19**

Not attempted

**3.20 Exercise 3.20**

Not attempted

**3.21 Exercise 3.21**

Not attempted

**3.22 Exercise 3.22**

Not attempted

**3.23 Exercise 3.23**

Not attempted

**3.24 Exercise 3.24**

Not attempted

**3.25 Exercise 3.25**

Not attempted

**3.26 Exercise 3.26**

Not attempted

**3.27 Exercise 3.27**

Not attempted

**3.28 Exercise 3.28**

Not attempted

**3.29 Exercise 3.29**

Not attempted

**3.30 Exercise 3.30**

Not attempted

**3.31 Exercise 3.31**

Not attempted

**3.32 Exercise 3.32**

Not attempted

**3.33 Exercise 3.33**

Not attempted

**3.34 Exercise 3.34**

Not attempted

**3.35 Exercise 3.35**

Not attempted

**3.36 Exercise 3.36**

Not attempted

**3.37 Exercise 3.37**

Deriving a maximum likelihood estimator for a histogram-like density model. Space  $\mathbf{x}$  is divided into fixed regions with density  $p(\mathbf{x}) = h_i \forall i$  regions. Volume  $i$  is denoted  $\Delta_i$ .  $N$  total observations



such that  $n_i$  observations are in region  $i$ .

$$h_i = \frac{n_i}{N\Delta_i} \quad (116)$$

$$\sum_i h_i \Delta_i = 1 \quad (117)$$

$$L = \prod_{i=0}^K h_i^{n_i} \quad (118)$$

$$\log L = \prod_{i=0}^K n_i \log h_i \quad (119)$$

$$\max \sum_{i=0}^K n_i \log h_i - \lambda \left( \sum_i h_i \Delta_i - 1 \right) \quad (120)$$

$$0 = \frac{n_i}{h_i} - \lambda \Delta_i \quad (121)$$

$$\frac{n_i}{h_i} = \lambda \Delta_i \quad (122)$$

$$h_i = \frac{n_i}{\lambda \Delta_i} \quad (123)$$

$$\text{Plugging into normalization:} \quad (124)$$

$$\sum_i \frac{n_i}{\lambda} = 1 \quad (125)$$

$$\lambda = N \quad (126)$$

$$h_i = \frac{n_i}{N\Delta_i} \quad (127)$$

### 3.38 Exercise 3.38

Not attempted