Homework 1

Jingwei Zhang 201528013229095

2015-10-15

1 Problem 1

1.1 a

Since:

$$P(error|x) = \begin{cases} P(\omega_2|x) & \text{if } x > \theta \\ P(\omega_1|x) & \text{if } x \le \theta \end{cases}$$

Then:

$$\begin{aligned} P(error) &= \int_{-\infty}^{+\infty} p(error, x)\ \mathrm{d}x \\ &= \int_{-\infty}^{\theta} p(\omega_1, x)\ \mathrm{d}x + \int_{\theta}^{+\infty} p(\omega_2, x)\ \mathrm{d}x \\ &= \int_{-\infty}^{\theta} p(x|\omega_1)P(\omega_1)\ \mathrm{d}x + \int_{\theta}^{+\infty} p(x|\omega_2)P(\omega_2)\ \mathrm{d}x \\ &= P(\omega_1)\int_{-\infty}^{\theta} p(x|\omega_1)\ \mathrm{d}x + P(\omega_2)\int_{\theta}^{+\infty} p(x|\omega_2)\ \mathrm{d}x \end{aligned}$$

1.2 b

Since x is defined on $(-\infty, +\infty)$, if $P(error)$ has minimum, $P'(error)$ must be 0.

$$\begin{aligned} P'(error) &= [P(\omega_1)\int_{-\infty}^{\theta} p(x|\omega_1)\ \mathrm{d}x + P(\omega_2)\int_{\theta}^{+\infty} p(x|\omega_2)\ \mathrm{d}x]' \\ &= P(\omega_1)[\int_{-\infty}^{\theta} p(x|\omega_1)\ \mathrm{d}x]' - P(\omega_2)[\int_{+\infty}^{\theta} p(x|\omega_2)\ \mathrm{d}x]' \\ &= P(\omega_1)p(\theta|\omega_1) - P(\omega_2)p(\theta|\omega_2) \\ &= 0 \end{aligned}$$

We can get:

$$P(\omega_1)p(\theta|\omega_1) = P(\omega_2)p(\theta|\omega_2)$$

1.3 c

No.$P'(error) = 0$ for specific x only guarantees that x is an stagnation point. It can be a local maximum, a local minimum or just nothing. And there may be multiple θ that satisfy the equation.

1.4 d

Suppose:

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}$$
$$P(x|\omega_1) = \frac{1}{\sqrt{2\pi}}e^{-(x+1)^2/2}$$
$$P(x|\omega_2) = \frac{1}{\sqrt{2\pi}}e^{-(x-1)^2/2}$$

Then:

$$P(error) = \frac{1}{2}\int_{-\infty}^{\theta} \frac{1}{\sqrt{2\pi}}e^{-(x+1)^2/2}\,\mathrm{d}x + \frac{1}{2}\int_{-\infty}^{\theta} \frac{1}{\sqrt{2\pi}}e^{-(x-1)^2/2}\,\mathrm{d}x$$
$$P'(error) = \frac{1}{2\sqrt{2\pi}}(e^{-(\theta+1)^2/2} + e^{-(\theta-1)^2/2})$$
$$P''(error) = \frac{-\theta}{\sqrt{2\pi}}(e^{-(\theta+1)^2/2} + e^{-(\theta-1)^2/2})$$

So, $P(\omega_1)p(\theta|\omega_1) = P(\omega_2)p(\theta|\omega_2)(P'(error) = 0)$ iff $\theta = 0$. And $P''(error) < 0, \forall\,\theta \in R$, which means that $P'(error) < 0, \forall\,\theta \in (-\infty, 0)$ and $P'(error) > 0, \forall\,\theta \in (0, +\infty)$. So that when $\theta = 0$,$P(error)$gets its global maximum.

2 Problem 3

2.1 a

Suppose $\mu_1 \leq \mu_2$:

From the probability function we can get that:

$$\begin{cases} p(x|\omega_1) \geq p(x|\omega_1), \text{if } x \leq \frac{\mu_2+\mu_1}{2} \\ p(x|\omega_1) < p(x|\omega_1), \text{if } x > \frac{\mu_2+\mu_1}{2} \end{cases}$$

So, we decide ω_1 if $x \leq (\mu_2 + \mu_1)/2$, otherwise decide ω_2, which minimize P_e. Then the probability of error would be:

$$\begin{aligned}
P_e &= \int_{-\infty}^{t} p(\omega_2, x)\,\mathrm{d}x + \int_{t}^{+\infty} p(\omega_1, x)\,\mathrm{d}x \quad (t = \frac{\mu_2 + \mu_1}{2}) \\
&= P(\omega_2)\int_{-\infty}^{t} p(x|\omega_2)\,\mathrm{d}x + P(\omega_1)\int_{t}^{+\infty} p(x|\omega_1)\,\mathrm{d}x \\
&= \frac{1}{2\sqrt{2\pi}\sigma}(\int_{-\infty}^{t} e^{-\frac{(x-\mu_2)^2}{2\sigma^2}}\,\mathrm{d}x + \int_{t}^{+\infty} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}}\,\mathrm{d}x) \\
&\text{Let } u_2 = \frac{x-\mu_2}{\sigma}, u_1 = \frac{x-\mu_1}{\sigma}: \\
&= \frac{1}{2\sqrt{2\pi}}(\int_{-\infty}^{\frac{\mu_1-\mu_2}{2}} e^{-\frac{u_2^2}{2}}\,\mathrm{d}u_2 + \int_{\frac{\mu_2-\mu_1}{2}}^{+\infty} e^{-\frac{u_1^2}{2}}\,\mathrm{d}u_1) \\
&\text{Let } u = u_1 = -u_2 \\
&= \frac{1}{2\sqrt{2\pi}}(\int_{\frac{\mu_2-\mu_1}{2}}^{+\infty} e^{-\frac{u^2}{2}}\,\mathrm{d}u + \int_{\frac{\mu_2-\mu_1}{2}}^{+\infty} e^{-\frac{u^2}{2}}\,\mathrm{d}u) \\
&= \frac{1}{\sqrt{2\pi}}\int_{a}^{+\infty} e^{-\frac{u^2}{2}}\,\mathrm{d}u \quad (a = \frac{\mu_2 - \mu_1}{\sigma})
\end{aligned}$$

When $\mu_1 > \mu_2$, the process is similar, we can get:

$$P_e = \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} e^{-\frac{u^2}{2}} \, \mathrm{d}u \quad (a = \frac{\mu_1 - \mu_2}{\sigma})$$

Above all, the equation$P_e = \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} e^{-\frac{u^2}{2}} \, \mathrm{d}u \quad (a = \frac{|\mu_1 - \mu_2|}{\sigma})$ can be proven.

2.2 b

Since:

$$\lim_{a\to\infty} e^{-a^2/2} = 0$$
$$\lim_{a\to\infty} \frac{1}{\sqrt{2\pi}a} = 0$$

We can get:

$$\begin{aligned}
\lim_{a\to\infty} P_e &= \lim_{a\to\infty} \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} e^{-\frac{u^2}{2}} \, \mathrm{d}t \\
&\le \lim_{a\to\infty} \frac{1}{\sqrt{2\pi}a} e^{-a^2/2} \\
&= \lim_{a\to\infty} e^{-a^2/2} \lim_{a\to\infty} \frac{1}{\sqrt{2\pi}a} \\
&= 0
\end{aligned}$$

3 Problem 4

3.1 a

Firstly we can get this probability dense table:

Table 1: Probability dense of $p(x_i|\omega_j)$

$p(x_i\|\omega_j)$	ω_1	ω_2	ω_3
x_1	0.3332	0.3970	0.3683
x_2	0.3970	0.3683	0.2661
x_3	0.2661	0.3683	0.3970
x_4	0.2179	0.3332	0.3970

Since these samples are independent, we can get:

$$\begin{aligned}
P(X|\omega) &= P(x_1|\omega_1)P(x_2|\omega_3)P(x3|\omega_3)P(x4|\omega_2) \\
&= 0.01173
\end{aligned}$$

$$\begin{aligned}
P(\omega) &= P(\omega_1)P(\omega_3)P(\omega_3)P(\omega_2) \\
&= \frac{1}{128}
\end{aligned}$$

$$
\begin{aligned}
P(X) &= \sum_{\omega_j} P(X, \omega_j) \\
&= \sum_{\omega_j} P(X|\omega_j)P(\omega_j) \\
&= \sum_{j_1=1}^{3}\sum_{j_2=1}^{3}\sum_{j_3=1}^{3}\sum_{j_4=1}^{3} P(x_1|\omega_{j_1})P(\omega_{j_1})P(x_2|\omega_{j_2})P(\omega_{j_2})P(x3|\omega_{j_3})P(\omega_{j_3})P(x4|\omega_{j_4})P(\omega_{j_4}) \\
&= 0.01208
\end{aligned}
$$

Then:

$$
\begin{aligned}
P(\omega|X) &= \frac{P(X|\omega)P(\omega)}{P(X)} \\
&= \frac{0.01173/128}{0.01208} \\
&= 0.007584
\end{aligned}
$$

3.2 b

Similarly:

$$
\begin{aligned}
P(X|\omega) &= P(x_1|\omega_1)P(x_2|\omega_2)P(x3|\omega_2)P(x4|\omega_3) \\
&= 0.01793
\end{aligned}
$$

$$
\begin{aligned}
P(\omega) &= P(\omega_1)P(\omega_2)P(\omega_2)P(\omega_3) \\
&= \frac{1}{128}
\end{aligned}
$$

$P(X)$ is exactly the same as above, Then:

$$
\begin{aligned}
P(\omega|X) &= \frac{P(X|\omega)P(\omega)}{P(X)} \\
&= \frac{0.01793/128}{0.01208} \\
&= 0.01160
\end{aligned}
$$

3.3 c

Since these points are independent, we will find the following:

$$
\begin{aligned}
\arg\max_{\omega} P(\omega|X) &= \arg\max_{\omega} \left(P(X|\omega)P(\omega)\frac{1}{P(X)}\right) \\
&= \arg\max_{\omega} \left(P(X|\omega)P(\omega)\right) \\
&= \arg\max_{\omega_i} \left(P(x_1|\omega_i)P(\omega_i)\right), \ldots, \arg\max_{\omega_i} \left(P(x_4|\omega_i)P(\omega_i)\right) \\
&= \omega_1, \omega_1, \omega_1, \omega_1 \quad \text{(It represents the sequence)}
\end{aligned}
$$

So the sequence is $\omega_1\omega_1\omega_1\omega_1$.

4 Problem 5

4.1 a

Since the samples are drawn by successive and independent selections of ω_i with probability $P(\omega_i)$,we have:

$$
P(z_{ik}|\omega_i) \sim B(1, P(\omega_i)) = P(\omega_i)^{z_{ik}}(1 - P(\omega_i))^{1-z_{ik}}
$$

Since these selections are independent,we have:

$$\begin{aligned}
P(z_{i1},\ldots,z_{in}|P(\omega_i)) &= \prod_{k=1}^{n} P(z_{ik}|P(\omega_i)) \\
&= \prod_{k=1}^{n} P(\omega_i)^{z_{ik}}(1-P(\omega_i))^{1-z_{ik}}
\end{aligned}$$

4.2 b

The log-ML function is:

$$\begin{aligned}
l(P(\omega_i)) &= \ln P(z_{i1},\ldots,z_{in}|P(\omega_i)) \\
&= \sum_{k=1}^{n}(z_{ik}\ln P(\omega_i) + (1-z_{ik})\ln(1-P(\omega_i))) \\
&= \ln P(\omega_i)\sum_{k=1}^{n} z_{ik} + \ln(1-P(\omega_i))(n-\sum_{k=1}^{n} z_{ik})
\end{aligned}$$

To maximize $l(P(\omega_i))$, let the derivative of l be 0:

$$\begin{aligned}
\frac{\mathrm{d}l(P(\omega_i))}{\mathrm{d}P(\omega_i)} &= \frac{\sum_{k=1}^{n} z_{ik}}{P(\omega_i)} - \frac{n-\sum_{k=1}^{n} z_{ik}}{1-P(\omega_i)} \\
&= \frac{\sum_{k=1}^{n} z_{ik} - nP(\omega_i)}{P(\omega_i)(1-P(\omega_i))} \\
&= 0
\end{aligned}$$

So, we have:

$$\sum_{k=1}^{n} z_{ik} - nP(\omega_i) = 0$$

Rewrite it, we have:

$$\hat{P}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n} z_{ik}$$

5 Problem 6

5.1 Key Code

```
# Problem a)
def gen_uniform(xl, xr, n):
    return np.random.randint(xl, xr + 1, n)

# Problem b)
def gen_paramaters():
    xl = np.random.randint(X_MIN, X_MAX+1)
    xr = np.random.randint(X_MIN, X_MAX+1)
    if xl > xr:
        xl, xr = xr, xl
    n = np.random.randint(N_MIN, N_MAX+1)
    return xl, xr, n

# Problem c) and d)
```

```python
def gen_rand_uniform(total_n):
    cnt = 0
    l = []
    while(cnt < total_n):
        xl, xr, n = gen_paramaters()
        if cnt + n > total_n:
            n = total_n - cnt
        l.extend(gen_uniform(xl, xr, n))
        cnt += n
    mean = np.mean(l)
    sigma = np.std(l)
    return l, mean, sigma
```

5.2 Results

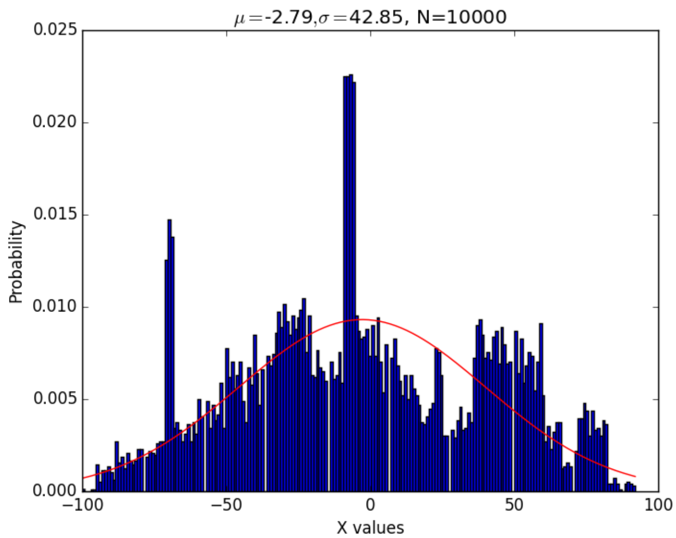

Figure 1: Figure for 10^4

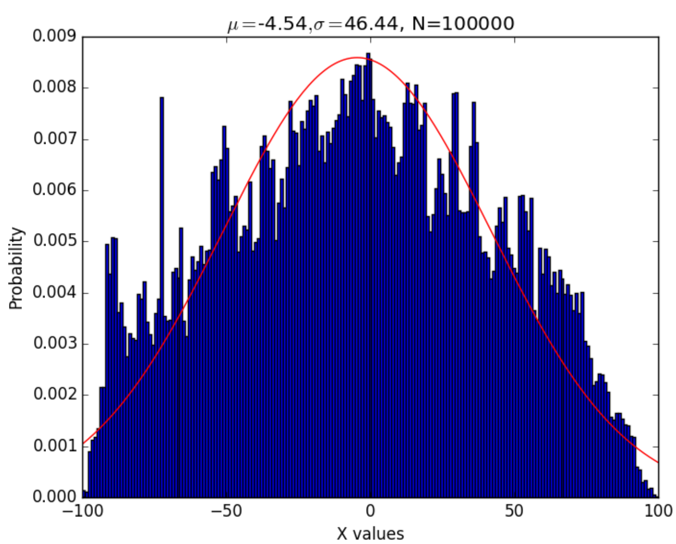

Figure 2: Figure for 10^5

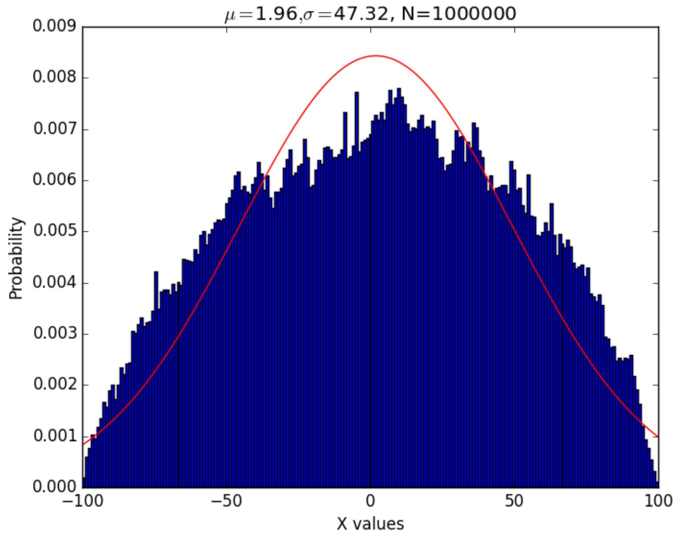

Figure 3: Figure for 10^6

5.3 e) Discuss

It can be seen from above three figures that when the size of sample is small(10^4), the approximation is not so good.However when the size of sample become larger,as to 10^5 and ^{6} it is very clear that independent random variables approximate a Gaussian.This illustrate the fact that the approximation can only be applied when the size of sample is sufficiently large.

6 Problem 8

6.1 Key Code

```python
# Problem a)
def calc_MLE_1D(sample, i):
    l = []
    for x in sample:
        l.append(x[i])
    mean = np.mean(l)
    sigma = np.std(l)
    return mean, sigma


def solve_a():
    l = [0, 1, 2]
    sample = samples[0]
    for j in l:
        mean, sigma = calc_MLE_1D(sample, j)
        print("\tomega1: X" + str(j + 1) +
              ": mu = " + str(mean) + ", sigma^2 = " + str(sigma**2))
    print()

# compute mean \u and covarient matrix \Sigma
def calc_MLE_Above_1D(l):
    mean = np.mean(l, axis=0)
    d = len(l[0])
    Sigma = np.zeros((d, d))
```

```
    for x in l:
        v = np.matrix([x - mean])
        Sigma += v.T.dot(v)
    Sigma /= len(l)
    return mean, Sigma

# Problem b)
def solve_b():
    for i in range(3):
        for j in range(i+1, 3):
            l = [[x[i],x[j]] for x in sample_1]
            mean, Sigma = calc_MLE_Above_1D(l)
            print("For X"+str(i+1)+" and X"+str(j+1)+":")
            print("mu =",mean)
            print("Sigma=\n",Sigma)
    print()

# Problem c)
def solve_c():
    mean, Sigma = calc_MLE_Above_1D(sample_1)
    print("mu =",mean)
    print("Sigma=\n",Sigma)

# Problem d)
def solve_d():
    l = [0, 1, 2]
    sample = samples[1]
    ans = []
    for j in l:
        mean, sigma = calc_MLE_1D(sample, j)
        ans.append([mean, sigma])
    mu = [mean for mean, sigma in ans]
    sigma_2 = [sigma*sigma for mean, sigma in ans]
    print("mu:", mu)
    print("sigma^2 1,2,3:",sigma_2)
    print()
```

6.2 Results

Problem a:

Table 2: Problem a: μ and σ^2 of x_i

x_i	μ	σ^2
x_1	-0.0709	0.90617729
x_2	-0.6047	4.20071481
x_3	-0.9110	4.54194900

Problem b:

For (x_1, x_2):

$$\mu_{12} = (-0.0709, -0.6047) \qquad \Sigma_{12} = \begin{bmatrix} 0.90617729 & 0.56778177 \\ 0.56778177 & 4.20071481 \end{bmatrix}$$

For (x_1, x_3):

$$\mu_{13} = (-0.0709, -0.9110) \qquad \Sigma_{13} = \begin{bmatrix} 0.90617729 & 0.3940801 \\ 0.3940801 & 4.5419490 \end{bmatrix}$$

For (x_2, x_3):

$$\mu_{23} = (-0.6047, -0.9110) \qquad \Sigma_{23} = \begin{bmatrix} 4.20071481 & 0.7337023 \\ 0.7337023 & 4.541949 \end{bmatrix}$$

Problem c:

$$\mu = (-0.0709, -0.6047, -0.911) \qquad \Sigma = \begin{bmatrix} 0.90617729 & 0.56778177 & 0.3940801 \\ 0.56778177 & 4.20071481 & 0.7337023 \\ 0.3940801 & 0.7337023 & 4.541949 \end{bmatrix}$$

Problem d:

$$\mu = (-0.1126, 0.4299, 0.003720)$$
$$\Sigma = diag(0.053925840, 0.045970090, 0.0072655056)$$

6.3 e and f

From result part we find that the corresponding μ and σ^2 of x_i are all the same in Problem a,b,c. While in the result of problem b and c, we know that $x1, x2, x3$ are not independent since Σ is not a diagonal matrix. This means that the dependency of x_i does not affect their mean and variance because their equations, $\hat{\mu_i} = \frac{1}{n}\sum_{i=1}^{n} x_i, \hat{Var}(x_i) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu_i})^2$, have nothing to do with other $x_j (j \neq i)$.

Comparing the result we get in problem a,b,c and problem d, their μ_i and σ_i^2 are different since they are estimated from different samples of different type ω.