Midterm

Q1. In a two-class problem, if the loss function is $\lambda_{11} = \lambda_{22} = 0$, $\lambda_{12} = 10$, and $\lambda_{21} = 5$.

a) Write the optimal decision rule

Solution. Let C_1, C_2 be the two given classes. Let x be the observed variable (or we can treat x as an input). Let us define actions α_1, α_2 as decisions to assign the input x to classes C_1, C_2, respectively. Then, the *expected risk* for taking action α_1 is

$$R(\alpha_1|x) = \lambda_{11}P(C_1|x) + \lambda_{12}P(C_2|x) = 10P(C_2|x) = 10(1 - P(C_1|x)) \tag{1}$$

The *expected risk* for taking action α_2 is

$$R(\alpha_2|x) = \lambda_{21}P(C_1|x) + \lambda_{22}P(C_2|x) = 5P(C_1|x) \tag{2}$$

To give the optimal decision rule, we select the class whose risk is minimal. In other words, the optimal decision rule is to

- choose C_1 as predicted output, if $R(\alpha_1|x) \leq R(\alpha_2|x)$
- choose C_2 as predicted output, if $R(\alpha_2|x) \leq R(\alpha_1|x)$

Combining with (1) and (2), this optimal decision rule becomes

- choose C_1 as predicted output, if $P(C_1|x) \geq 2/3$
- choose C_2 as predicted output, if $P(C_1|x) \leq 2/3$

■

b) How does the rule change if we add a third action of reject with $\lambda = 1$?

Solution. Let us define α_3 as the action when we neither assign the input x to C_1 nor C_2 (so-called *reject*). Then the *expected risk* of this "reject" action is

$$R(\alpha_3|x) = \lambda P(C_1|x) + \lambda P(C_2|x) = \lambda = 1 \qquad \text{(since } P(C_1|x) + P(C_2|x) = 1)$$

Since, we introduce a new action, the optimal decision rule changes. In particular, we

- choose C_1 if **both** $R(\alpha_1|x) \leq R(\alpha_2|x)$ **and** $R(\alpha_1|x) \leq R(\alpha_3|x)$
- choose C_2 if **both** $R(\alpha_2|x) \leq R(\alpha_1|x)$ **and** $R(\alpha_2|x) \leq R(\alpha_3|x)$
- reject if $R(\alpha_3|x) \leq both\{R(\alpha_1|x), R(\alpha_2|x)\}$

Combining with (1) and (2), this rule can be rewritten as

- choose C_1 if $P(C_1|x) \geq 9/10$
- choose C_2 if $P(C_1|x) \leq 1/5$
- reject if $1/5 \leq P(C_1|x) \leq 9/10$

■

Q2. Apply candidate elimination algorithm on following dataset to compute version space if there is any. Show all your steps.

Size	**Color**	**Shape**	**Class**
Big	Red	Circle	-ve
Small	Red	Triangle	-ve
Small	Red	Circle	+ve
Big	Blue	Circle	-ve
Small	Blue	Circle	+ve

Solution. Let S_i, G_i be the most specific hypothesis and most general hypothesis in the loop i. We have

Loop #	**Current** S	**Current** G
Initialize	$S_0 = (\varnothing, \varnothing, \varnothing)$	$G_0 = (?, ?, ?)$
1	$S_1 = (\varnothing, \varnothing, \varnothing)$	$G_1 = (?, ?, ?), (S, ?, ?), (?, B, ?), (?, ?, T)$
2	$S_2 = (\varnothing, \varnothing, \varnothing)$	$G_2 = (S, ?, ?), (?, B, ?), (?, ?, T), (S, B, ?), (S, ?, C), (B, ?, T), (?, B, T)$
3	$S_3 = (S, R, C)$	$G_3 = \cancel{(?, B, ?)}, \cancel{(S, B, ?)}, (S, ?, C), \cancel{(B, ?, T)}, \cancel{(?, B, T)}$
4	$S_4 = (S, R, C)$	$G_4 = (S, ?, C)$
5	$S_5 = (S, ?, C)$	$G_5 = (S, ?, C)$

Therefore, the only one possible version space is $(S, ?, C)$.

∎

Q3. What is the likelihood ratio

$$\frac{p(x|C_1|)}{p(x|C_2|)}$$

in the case of Gaussian densities?

Q4. Apply the Apriori algorithm to find the frequent itemset of size 2 and 3. Also, write the association rule if possible. Assume support = 3 and confidence = 60%.

Transaction No.	**Items**
T1	1, 2, 3, 4, 5, 6
T2	7, 2, 3, 4, 5, 6
T3	1, 8, 4, 5
T4	1, 9, 0, 4, 6
T5	0, 2, 2, 4, 5

Solution. The following are the steps:

1-item sets	**Frequency**	**Accepted?**
$\{0\}$	2	No
$\{1\}$	3	Yes
$\{2\}$	3	Yes
$\{3\}$	2	No
$\{4\}$	5	Yes
$\{5\}$	4	Yes
$\{6\}$	3	Yes
$\{7\}$	1	No
$\{8\}$	1	No
$\{9\}$	1	No

Table 1: 1-item sets contain distinct elements $\{1, 2, 4, 5, 6\}$

2-item sets	**Frequency**	**Accepted?**
$\{1, 2\}$	1	No
$\{1, 4\}$	3	Yes
$\{1, 5\}$	2	No
$\{1, 6\}$	2	No
$\{2, 4\}$	3	Yes
$\{2, 5\}$	3	Yes
$\{2, 6\}$	2	No
$\{4, 5\}$	4	Yes
$\{4, 6\}$	3	Yes
$\{5, 6\}$	2	No

Table 2: 2-item sets contain distinct elements $\{1, 2, 4, 5, 6\}$

3-item sets	**Frequency**	**Accepted?**
$\{1, 2, 4\}$	0	No
$\{1, 4, 5\}$	2	No
$\{1, 4, 6\}$	2	No
$\{2, 4, 5\}$	3	Yes
$\{2, 4, 6\}$	2	No
$\{1, 2, 5\}$	1	No
$\{2, 5, 6\}$	2	No
$\{4, 5, 6\}$	2	No

Table 3: 3-item sets contain distinct elements $\{1, 2, 4, 5, 6\}$

$$Confidence(1 \to 4) = \frac{3}{3} = 100\%$$
$$Confidence(4 \to 1) = \frac{3}{5} = 60\%$$
$$Confidence(2 \to 4) = \frac{3}{3} = 100\%$$
$$Confidence(4 \to 2) = \frac{3}{5} = 60\%$$
$$Confidence(2 \to 5) = \frac{3}{3} = 100\%$$

$$Confidence(5 \to 2) = \frac{3}{4} = 75\%$$
$$Confidence(4 \to 5) = \frac{4}{5} = 80\%$$
$$Confidence(5 \to 4) = \frac{4}{4} = 100\%$$
$$Confidence(4 \to 6) = \frac{3}{5} = 60\%$$
$$Confidence(6 \to 4) = \frac{3}{3} = 100\%$$

$$Confidence(2 \to 4, 5) = \frac{3}{3} = 100\%$$
$$Confidence(4 \to 2, 5) = \frac{3}{5} = 60\%$$
$$Confidence(5 \to 2, 4) = \frac{3}{4} = 75\%$$
$$Confidence(2, 4 \to 5) = \frac{3}{3} = 100\%$$
$$Confidence(2, 5 \to 4) = \frac{3}{3} = 100\%$$
$$Confidence(4, 5 \to 2) = \frac{3}{4} = 75\%$$

■

Q5. We have a test which is effective in that if a person has the disease, there is a 99% chance that the test result is positive. There is a 1 in 1000 chance that the test result is positive in the case of a healthy person. The disease is seen in 1 out of a million people. Assume that a new person arrives and the test result is positive. What is the probability that the patient has the disease?

Solution. Let us define

- D be the event that a person has a disease.
- $+$ be the event that the test result is positive.

From the given assumption, we get

$$\begin{aligned} P(D) &= \frac{1}{1,000,000} = 0.00001 \\ P(+|D) &= \frac{99}{100} = 0.99 \\ P(+|\overline{D}) &= \frac{1}{1,000} = 0.001 \end{aligned}$$

By applying the Bayesian Rule, the probability that a new person has a disease when the test result is positive is calculated by

$$\begin{aligned} P(D|+) &= \frac{P(D)P(+|D)}{P(+)} && \text{(Bayesian rule)} \\ &= \frac{P(D)P(+|D)}{P(+\ ,\ D) + P(+\ ,\ \overline{D})} && \text{(Law of total probability)} \\ &= \frac{P(D)P(+|D)}{P(+|D)P(D) + P(+|\overline{D})P(\overline{D})} \\ &= \frac{0.00001 \times 0.99}{0.99 \times 0.00001 + 0.001 \times (1 - 0.00001)} \\ &\approx 0.98\% \end{aligned}$$

■