Machine Learning Module:

Bayesian Learning Problem Set Answer

Q1.

Given information:

$P(W) = 0.7$

$P(L) = 0.3$

$P(P|W) = 0.9$

$P(P|L) = 0.6$

Using Bayes' theorem: $P(W|P) = [P(P|W) \times P(W)] \div P(P)$

To find $P(P)$, we use the law of total probability:

$P(P) = P(P|W) \times P(W) + P(P|L) \times P(L)$ $P(P)$

$= 0.9 \times 0.7 + 0.6 \times 0.3$ $P(P)$

$= 0.63 + 0.18$ $P(P) = 0.81$

Now we can calculate $P(W|P)$:

$P(W|P) = [P(P|W) \times P(W)] \div P(P)$ $P(W|P)$

$= (0.9 \times 0.7) \div 0.81$ $P(W|P) = 0.63 \div 0.81$

$P(W|P) = 0.7778 \approx 0.778$ or 77.8%

So, the probability that Manchester United won the game is approximately 0.778 or 77.8%.

Q2.

Given,

$P(F) = 0.3$

$P(R) = 0.7$

$P(D|F) = 0.8$

$P(D|R) = 0.1$

Using Bayes' Rule: $P(F|D) = [P(D|F) \times P(F)] \div P(D)$

We use the law of total probability:

$P(D) = P(D|F) \times P(F) + P(D|R) \times P(R)$ $P(D)$

$= 0.8 \times 0.3 + 0.1 \times 0.7$

$P(D) = 0.24 + 0.07$ $P(D) = 0.31$

Now we can calculate $P(F|D)$:

$P(F|D) = [P(D|F) \times P(F)] \div P(D)$ $P(F|D)$

$= (0.8 \times 0.3) \div 0.31$ $P(F|D) = 0.24 \div 0.31$

$P(F|D) = 0.7742 \approx 0.774$

Therefore, given that Mr. Smith died, the probability that the nurse forgot to give him the pill is approximately 0.774 or 77.4%.

Q3.

Given:

$P(G) = 0.1$

$P(C) = 0.3$

$P(N) = 0.6$

$P(T|G) = 0.8$

$P(T|C) = 0.4$

$P(T|N) = 0.2$

Using Bayes' theorem: $P(G|T) = [P(T|G) \times P(G)] \div P(T)$

To find $P(T)$, I'll use the law of total probability:

$P(T) = P(T|G) \times P(G) + P(T|C) \times P(C) + P(T|N) \times P(N)$

$P(T) = 0.8 \times 0.1 + 0.4 \times 0.3 + 0.2 \times 0.6$

$P(T) = 0.08 + 0.12 + 0.12$ $P(T) = 0.32$

Now calculate $P(G|T)$:

$P(G|T) = [P(T|G) \times P(G)] \div P(T)$ $P(G|T)$

$= (0.8 \times 0.1) \div 0.32$ $P(G|T) = 0.08 \div 0.32$ Therefore, $P(G|T) = 0.25$

Therefore, given a positive test result, the probability of gold being found on the University of Warwick campus is 0.25 or 25%.

Q4.

Given:

$P(M) = 0.05$

$P(M') = 0.95$

$P(T+|M) = 0.95$

$P(T-|M') = 0.70$

$P(T+|M') = 0.30$

Using Bayes' Rule: $P(M|T+) = [P(T+|M) \times P(M)] \div P(T+)$

To find $P(T+)$:

$P(T+) = P(T+|M) \times P(M) + P(T+|M') \times P(M')$

$P(T+) = 0.95 \times 0.05 + 0.30 \times 0.95$ $P(T+) = 0.0475 + 0.285$

$P(T+) = 0.3325$

Now calculating $P(M|T+)$: $P(M|T+)$

$= (0.95 \times 0.05) \div 0.3325$

$P(M|T+) = 0.0475 \div 0.3325$

$P(M|T+) = 0.1429 \approx 0.143$ or 14.3%

Therefore, given a positive test result.

For the second test,

Using Bayes' Rule again: $P(M|T+, T-) = [P(T-|M) \times P(M|T+)] \div P(T-|T+)$

$P(M|T+) = 0.143$

$P(T-|M) = 0.05$

$P(T-|M') = 0.70$

To find $P(T-|T+)$: $P(T-|T+)$

$= P(T-|M) \times P(M|T+) + P(T-|M') \times P(M'|T+)$

$P(T-|T+) = 0.05 \times 0.143 + 0.70 \times (1-0.143)$

$P(T-|T+) = 0.00715 + 0.70 \times 0.857$

$P(T-|T+) = 0.00715 + 0.5999$ $P(T-|T+) = 0.60705$

Now calculating $P(M|T+, T-)$:

$P(M|T+, T-) = (0.05 \times 0.143) \div 0.60705$

$= 0.00715 \div 0.60705 = 0.01178 \approx 0.012$ or 1.2%

Therefore, after a positive first test followed by a negative second test, the probability that the patient has Meningitis drops significantly from 14.3% to approximately 1.2%.

Q5.

Given:

$P(R) = 0.8$

$P(R') = 0.2$

$P(N|R) = 0.25$

$P(N|R') = 0.85$

Using Bayes' theorem: $P(R|N) = [P(N|R) \times P(R)] \div P(N)$

To find $P(N)$:

$P(N) = P(N|R) \times P(R) + P(N|R') \times P(R')$

$= 0.25 \times 0.8 + 0.85 \times 0.2 = 0.2 + 0.17 = 0.37$

Now calculating $P(R|N)$:

$P(R|N) = (0.25 \times 0.8) \div 0.37$

$$P(R|N) = 0.2 \div 0.37 \quad P(R|N) = 0.5405... \approx 0.54 \text{ or } 54\%$$

Therefore, given a negative result from the experiment, there's still approximately a 54% probability that it will rain the next day.

Q6.

Given information:

$P(C) = 0.0001$

$P(C') = 0.9999$

$P(J|C) = 0.64$

$P(J|C') = 0.6$

$P(T+|C) = 0.99$

$P(T+|C') = 0.04$

$P(J) = P(J|C) \times P(C) + P(J|C') \times P(C')$

$P(J) = 0.64 \times 0.0001 + 0.6 \times 0.9999 = 0.000064 + 0.59994 = 0.600004$

$P(C|J) = (0.64 \times 0.0001) \div 0.600004 = 0.000064 \div 0.600004$

$= 0.000107 \approx 0.000107$

Now, using P(C|J) as the new prior and incorporating the positive test:

$P(C|J, T+) = [P(T+|C) \times P(C|J)] \div P(T+|J)$

$P(T+|J) = P(T+|C) \times P(C|J) + P(T+|C') \times P(C'|J)$

$P(T+|J) = 0.99 \times 0.000107 + 0.04 \times (1-0.000107)$

$= 0.000106 + 0.04 \times 0.999893 = 0.000106 + 0.039996 = 0.040102$

$P(C|J, T+) = (0.99 \times 0.000107) \div 0.040102$

$= 0.000106 \div 0.040102 = 0.002642... \approx 0.0026 \text{ or } 0.26\%$

So, after the first positive test, the probability Fred has Chickungunya is only about 0.26%.

$P(T+|J, T+) = P(T+|C) \times P(C|J, T+) + P(T+|C') \times P(C'|J, T+)$

$P(T+|J, T+) = 0.99 \times 0.0026 + 0.04 \times (1-0.0026)$

$= 0.002574 + 0.04 \times 0.9974 = 0.002574 + 0.039896 = 0.04247$

$P(C|J, T+, T+) = (0.99 \times 0.0026) \div 0.04247$

$= 0.002574 \div 0.04247 = 0.06061... \approx 0.061$ or 6.1%

After the second positive test, the probability increases to about 6.1%, which is still quite low.

Let's continue this process iteratively. After n tests, we want to find the minimum value of n where $P(C|J, T+_1, T+..,T+_n) > 0.5$

For n=3:

$= 0.99 \times 0.061 + 0.04 \times 0.939 = 0.060 + 0.038 = 0.098$

$= (0.99 \times 0.061) \div 0.098 = 0.060 \div 0.098 = 0.612$ or 61.2%

After 3 positive tests, the probability exceeds 50% (reaching about 61.2%).

Q7.

Total records: 10

Churned "Yes": 4 records

Churned "No": 6 records

Handset = Old, Time Since Customer > 2.5 years, Age = 55

First, I'll calculate the prior probabilities:

$P(Yes) = 4/10 = 0.4$

$P(No) = 6/10 = 0.6$

Now, I'll calculate the conditional probabilities for each attribute:

For **Churn = Yes**:

$P(Old \mid Yes) = 2/4 = 0.5$

$P(> 2.5 \mid Yes) = 3/4 = 0.75$

$P(\geq 55 \mid Yes) = 2/4 = 0.5$

For **Churn = No**:

$P(\text{Old} \mid \text{No}) = 0/6 = 0$

$P(> 2.5 \mid \text{No}) = 4/6 \approx 0.67$

$P(\geq 55 \mid \text{No}) = 3/6 = 0.5$

Using the formula: $P(x|C) = (nx + m \times p)/ (n + m)$ Where:

Let's calculate with $m = 1$:

For **Churn = Yes** ($n = 4$):

$P(\text{Old} \mid \text{Yes}) = (2 + 1 \times 0.5)/ (4 + 1) = 2.5/5 = 0.5$

$P(> 2.5 \mid \text{Yes}) = (3 + 1 \times 0.5)/ (4 + 1) = 3.5/5 = 0.7$

$P(\geq 55 \mid \text{Yes}) = (2 + 1 \times 0.5)/ (4 + 1) = 2.5/5 = 0.5$

For **Churn = No** ($n = 6$):

$P(\text{Old} \mid \text{No}) = (0 + 1 \times 0.5)/ (6 + 1) = 0.5/7 \approx 0.071$

$P(> 2.5 \mid \text{No}) = (4 + 1 \times 0.5)/ (6 + 1) = 4.5/7 \approx 0.643$

$P(\geq 55 \mid \text{No}) = (3 + 1 \times 0.5)/ (6 + 1) = 3.5/7 = 0.5$

$P(\text{Churn} = \text{Yes} \mid X) \propto P(\text{Yes}) \times P(\text{Handset}$

$= \text{Old} \mid \text{Yes}) \times P(\text{Time} > 2.5 \mid \text{Yes}) \times P(\text{Age} \geq 55 \mid \text{Yes})$

$P(\text{Churn} = \text{Yes} \mid X) \propto 0.4 \times 0.5 \times 0.7 \times 0.5 = 0.07$

$P(\text{Churn} = \text{No} \mid X) \propto P(\text{No}) \times P(\text{Handset} = \text{Old} \mid \text{No}) \times P(\text{Time} > 2.5 \mid \text{No}) \times P(\text{Age} \geq 55 \mid \text{No})$ $P(\text{Churn} = \text{No} \mid X) \propto 0.6 \times 0.071 \times 0.643 \times 0.5 \approx 0.014$

Total = 0.07 + 0.014 = 0.084

$P(\text{Churn} = \text{Yes} \mid X) = 0.07/0.084 \approx 0.833$ or 83.3%

Therefore, $P(\text{Churn} = \text{No} \mid X) = 0.014/0.084 \approx 0.167$ or 16.7%

Q8.

Given:

$P(\text{Class} = 1) = 5/7 \approx 0.714$

$P(\text{Class} = 0) = 2/7 \approx 0.286$

For Gaussian distributions, I need to calculate the mean (μ) and variance (σ^2) for each feature in each class.

For Class = 1:

X values: [5, 5, 3, 4, 3]

- Mean $\mu_{x1} = (5+5+3+4+3)/5 = 4.0$
- Variance $\sigma^2_{x1} = [(5-4)^2 + (5-4)^2 + (3-4)^2 + (4-4)^2 + (3-4)^2]/5 = 0.8$

Y values: [6, 7, 6, 5, 7]

- Mean $\mu y_1 = (6+7+6+5+7)/5 = 6.2$
- Variance $\sigma^2 y_1 = [(6-6.2)^2 + (7-6.2)^2 + (6-6.2)^2 + (5-6.2)^2 + (7-6.2)^2]/5 = 0.56$

For Class = 0:

X values: [2, 8]

- Mean $\mu_{x0} = (2+8)/2 = 5.0$
- Variance $\sigma^2_{x0} = [(2-5)^2 + (8-5)^2]/2 = 9.0$

Y values: [4, 6]

- Mean $\mu y_0 = (4+6)/2 = 5.0$
- Variance $\sigma^2 y_0 = [(4-5)^2 + (6-5)^2]/2 = 1.0$

Using the Gaussian probability density function:

$$f(x; \mu, \sigma^2) = (1/\sqrt{(2\pi\sigma^2)}) \times e^{(-(x-\mu)^2/(2\sigma^2))}$$

For customer with X = 7, Y = 4:

For Class = 1:

$P(X=7|Class=1) = (1/\sqrt{(2\pi\times0.8)}) \times e^{(-(7-4)^2/(2\times0.8))} = 0.446 \times e^{(-5.625)} = 0.446 \times 0.0036 \approx 0.0016$

$P(Y=4|Class=1) = (1/\sqrt{(2\pi\times0.56)}) \times e^{(-(4-6.2)^2/(2\times0.56))} = 0.534 \times e^{(-4.321)} = 0.534 \times 0.0133 \approx 0.0071$

For Class = 0:

$P(X=7|Class=0) = (1/\sqrt{(2\pi\times9)}) \times e^{(-(7-5)^2/(2\times9))} = 0.133 \times e^{(-0.222)} = 0.133 \times 0.801 \approx 0.1065$

$P(Y=4|Class=0) = (1/\sqrt{(2\pi\times1)}) \times e^{(-(4-5)^2/(2\times1))} = 0.399 \times e^{(-0.5)} = 0.399 \times 0.607 \approx 0.241$

Using Naive Bayes formula:

$$P(Class|X,Y) \propto P(Class) \times P(X|Class) \times P(Y|Class)$$

For Class = 1: $P(Class=1|X=7, Y=4) \propto 0.714 \times 0.0016 \times 0.0071 \approx 0.714 \times 1.14 \times 10^{(-5)} \approx 8.14 \times 10^{(-6)}$

For Class = 0: $P(Class=0|X=7, Y=4) \propto 0.286 \times 0.1065 \times 0.241 \approx 0.286 \times 0.0257 \approx 0.00733$

$P(Class=1|X=7, Y=4) = (8.14 \times 10^{(-6)})/((8.14 \times 10^{(-6)} + 0.00733)) \approx (8.14 \times 10^{(-6)})/(0.00734) \approx 0.0011$ or 0.11%

$P(Class=0|X=7, Y=4) = 0.00733/(0.00734) \approx 0.9989$ or 99.89%

The probability that the customer with X = 7 and Y = 4 is a positive example (Class = 1) is approximately 0.11%.

Q9.

For x = 2.8:

$f(2.8; 2.3, 1.8) = (1/(1.8\sqrt{(2\pi)})) * \exp(-(2.8-2.3)^2/(2*1.8^2)) = 0.2213 * \exp(-0.0772) = 0.2213 * 0.9257 \approx 0.2049$

$f(2.8; 6.8, 2.2) = (1/(2.2\sqrt{(2\pi)})) * \exp(-(2.8-6.8)^2/(2*2.2^2)) = 0.1811 * \exp(-4.1322) = 0.1811 * 0.0161 \approx 0.0029$

$p(2.8) = 0.5 * 0.2049 + 0.5 * 0.0029 \approx 0.1039$

For x = 4.2:

$f(4.2; 2.3, 1.8) = (1/(1.8\sqrt{(2\pi)})) * \exp(-(4.2-2.3)^2/(2*1.8^2)) = 0.2213 * \exp(-0.5556) = 0.2213 * 0.5738 \approx 0.1270$

$f(4.2; 6.8, 2.2) = (1/(2.2\sqrt{(2\pi)})) * \exp(-(4.2-6.8)^2/(2*2.2^2)) = 0.1811 * \exp(-0.7025) = 0.1811 * 0.4954 \approx 0.0897$

$p(4.2) = 0.5 * 0.1270 + 0.5 * 0.0897 \approx 0.1084$

For x = 5.3:

$f(5.3; 2.3, 1.8) = (1/(1.8\sqrt{(2\pi)})) * \exp(-(5.3-2.3)^2/(2*1.8^2)) = 0.2213 * \exp(-1.3889) = 0.2213 * 0.2493 \approx 0.0552$

$f(5.3; 6.8, 2.2) = (1/(2.2\sqrt{(2\pi)})) * \exp(-(5.3-6.8)^2/(2*2.2^2)) = 0.1811 * \exp(-0.2314) = 0.1811 * 0.7934 \approx 0.1437$

$p(5.3) = 0.5 * 0.0552 + 0.5 * 0.1437 \approx 0.0995$

For $x = 5.5$:

$f(5.5; 2.3, 1.8) = (1/(1.8\sqrt{(2\pi)})) * \exp(-(5.5-2.3)^2/(2*1.8^2)) = 0.2213 * \exp(-1.5741) = 0.2213 * 0.2072 \approx 0.0459$

$f(5.5; 6.8, 2.2) = (1/(2.2\sqrt{(2\pi)})) * \exp(-(5.5-6.8)^2/(2*2.2^2)) = 0.1811 * \exp(-0.1736) = 0.1811 * 0.8407 \approx 0.1522$

$p(5.5) = 0.5 * 0.0459 + 0.5 * 0.1522 \approx 0.0991$

For $x = 2.1$:

$f(2.1; 2.3, 1.8) = (1/(1.8\sqrt{(2\pi)})) * \exp(-(2.1-2.3)^2/(2*1.8^2)) = 0.2213 * \exp(-0.0123) = 0.2213 * 0.9877 \approx 0.2186$

$f(2.1; 6.8, 2.2) = (1/(2.2\sqrt{(2\pi)})) * \exp(-(2.1-6.8)^2/(2*2.2^2)) = 0.1811 * \exp(-4.6128) = 0.1811 * 0.0099 \approx 0.0018$

$p(2.1) = 0.5 * 0.2186 + 0.5 * 0.0018 \approx 0.1102$

Now I calculate the log-likelihood:

$L = \log(0.1039) + \log(0.1084) + \log(0.0995) + \log(0.0991) + \log(0.1102)$ $L = -2.2643 - 2.2220 - 2.3077 - 2.3117 - 2.2053$ $L \approx -11.3110$

Therefore, the log-likelihood of the data given the specified mixture model is approximately -11.31.

Start with initial parameter values:

$\mu_1 = 2.3, \sigma_1 = 1.8, \mu_2 = 6.8, \sigma_2 = 2.2, w_1 = w_2 = 0.5$

$\gamma_{11} = (0.5 * 0.2049) / 0.1039 \approx 0.9860$

$\gamma_{12} = (0.5 * 0.0029) / 0.1039 \approx 0.0140$

For $x = 4.2$:

$\gamma_{21} = (0.5 * 0.1270) / 0.1084 \approx 0.5859$

$\gamma_{22} = (0.5 * 0.0897) / 0.1084 \approx 0.4141$

For $x = 5.3$:

$\gamma_{31} = (0.5 * 0.0552) / 0.0995 \approx 0.2773$

$\gamma_{32} = (0.5 * 0.1437) / 0.0995 \approx 0.7227$

For $x = 5.5$:

$\gamma_{41} = (0.5 * 0.0459) / 0.0991 \approx 0.2314$

$\gamma_{42} = (0.5 * 0.1522) / 0.0991 \approx 0.7686$

For $x = 2.1$:

$\gamma_{51} = (0.5 * 0.2186) / 0.1102 \approx 0.9918$

$\gamma_{52} = (0.5 * 0.0018) / 0.1102 \approx 0.0082$

Therefore, an improved estimate for the mixture model parameters would be:

Component 1: $\mu_1 \approx 3.41$, $\sigma_1 \approx 1.15$, $w_1 \approx 0.61$

Component 2: $\mu_2 \approx 5.10$, $\sigma_2 \approx 0.58$, $w_2 \approx 0.39$