Universidade Federal do Rio de Janeiro
Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia

Programa de Engenharia de Sistemas e Computação

CPS863 - Aprendizado de Máquina
Prof. Dr. Edmundo de Souza e Silva
(PESC/COPPE/UFRJ)

Lista de Exercícios 1a

Luiz Henrique Souza Caldas
email: lhscaldas@cos.ufrj.br

16 de outubro de 2024

Questão 1

(Recordação)
Uma caixa contém três moedas: duas são normais e uma moeda falsa com duas caras (P(Ca)=1). Se você pegar uma moeda da caixa e jogá-la, qual a probabilidade de sair cara? Se você pegar uma moeda da caixa e jogá-la, e sair cara, qual a probabilidade de ser a moeda falsa?

Resposta:

1. Probabilidade de sair cara
Temos três moedas na caixa:

- Moeda 1 (M1): normal, $P(C|M_1) = 0.5$
- Moeda 2 (M2): normal, $P(C|M_2) = 0.5$
- Moeda 3 (M3): falsa, $P(C|M_3) = 1$

A probabilidade de escolher cada moeda é igual: $P(M_i) = \frac{1}{3}$.
A probabilidade total de sair cara $P(C)$ é:

$$P(C) = P(C|M_1) \cdot P(M_1) + P(C|M_2) \cdot P(M_2) + P(C|M_3) \cdot P(M_3)$$

Substituindo os valores:

$$P(C) = 0.5 \cdot \frac{1}{3} + 0.5 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}$$

Assim, a probabilidade de sair cara é $P(C) = \frac{2}{3}$.
2. Probabilidade de ser a moeda falsa, dado que saiu cara
Para a probabilidade de que a moeda escolhida seja a falsa, dado que saiu cara, usamos o Teorema de Bayes:

$$P(M_3|C) = \frac{P(C|M_3) \cdot P(M_3)}{P(C)} = \frac{1 \cdot \frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

Questão 2

(Material introdutório)
Uma urna U_A tem $N = 1000$ bolas sendo 25% delas azuis e o restante pretas. Uma outra urna U_B também contém $N = 1000$ bolas, mas apenas 10% delas são azuis (e o restante pretas). As urnas são idênticas externamente, exceto por uma marcação, U_A, U_B, que permite a identificação de cada uma. Entretanto, essa identificação está na parte inferior das urnas, de forma que não é possível visualizar o rótulo, exceto se a urna for levantada.

- João tira (de olhos vendados) 2 bolas azuis de uma das urnas. Você vai ter que adivinhar a urna escolhida. Se a probabilidade de João escolher uma das urnas for a mesma, qual a aposta que você fará? Note que, para fazer a aposta, você precisa determinar qual a probabilidade das bolas serem provenientes da urna U_A. Você tem confiança na sua aposta? Por que?

- Um amigo seu diz que João sabe a posição das urnas e escolhe a urna U_A com probabilidade 0.15. Sua aposta mudaria? Você teria confiança na sua aposta? Justifique a resposta.

Resposta:

1. Aposta com probabilidade de escolha de urna $P[U_A] = 0.5$:
As probabilidades de tirar 2 bolas azuis são:

$$P(2 \text{ azuis}|U_A) = \frac{250}{1000} \cdot \frac{249}{999} = \frac{249}{3996} \approx 0.0623$$

$$P(2 \text{ azuis}|U_B) = \frac{100}{1000} \cdot \frac{99}{999} = \frac{9.9}{999} \approx 0.0099$$

Aplicando o Teorema de Bayes, podemos calcular a probabilidade de que as bolas azuis tenham vindo de cada urna, dado que a probabilidade de escolha de urna é a mesma:

$$P(U_A|2 \text{ azuis}) = \frac{P(2 \text{ azuis}|U_A) \cdot 0.5}{P(2 \text{ azuis}|U_A) \cdot 0.5 + P(2 \text{ azuis}|U_B) \cdot 0.5} \approx 0.862$$

$$P(U_B|2 \text{ azuis}) = \frac{P(2 \text{ azuis}|U_B) \cdot 0.5}{P(2 \text{ azuis}|U_A) \cdot 0.5 + P(2 \text{ azuis}|U_B) \cdot 0.5} \approx 0.138$$

Assim, a aposta é que as bolas vieram da urna U_A, com uma probabilidade de 86.2%. A confiança na aposta é alta, pois a probabilidade de que as bolas vieram da urna U_B é de 13.8%.

2. Aposta com probabilidade de escolha de urna $P[U_A] = 0.15$:
Reaplicando o Teorema de Bayes, podemos calcular a probabilidade de que as bolas azuis tenham vindo de cada urna, dado que a probabilidade de escolha de urna é diferente ($P[U_A] = 0.15$ e $P[U_B] = 0.85$):

$$P(U_A|2 \text{ azuis}) = \frac{P(2 \text{ azuis}|U_A) \cdot 0.15}{P(2 \text{ azuis}|U_A) \cdot 0.15 + P(2 \text{ azuis}|U_B) \cdot 0.85} \approx 0.526$$

$$P(U_B|2 \text{ azuis}) = \frac{P(2 \text{ azuis}|U_B) \cdot 0.85}{P(2 \text{ azuis}|U_A) \cdot 0.15 + P(2 \text{ azuis}|U_B) \cdot 0.15} \approx 0.474$$

Assim, a aposta é que as bolas vieram da urna U_A, com uma probabilidade de 52.6%. A confiança na aposta é menor, pois a probabilidade de que as bolas vieram da urna U_B agora é de 47.4%.

Questão 3

Considere um dataset cujas amostras são obtidas independentemente a partir de uma distribuição discreta uniforme $U(1, 5)$. Considere um dataset com as seguintes amostras: $\{2, 2, 4, 3, 2\}$.

1. Qual a verossimilhança (likelihood) de observar essas amostras?

2. E o log-likelihood?

Resposta:

1. likelihood:
Seja:

- X a variável aleatória que representa a amostra.
- U a distribuição uniforme discreta $U(1, 5)$.
- $D = \{2, 2, 4, 3, 2\}$ o dataset.

A verossimilhança de observar as amostras é dada por:

$$L(U|D) = P(D|U) = \prod_{i=1}^{5} P(x_i|U)$$

Como a distribuição é uniforme, a probabilidade de cada amostra é $P(x_i|U) = \frac{1}{5}$. Assim, a verossimilhança é:

$$L(U|D) = P(2)^3 \cdot P(3)^1 \cdot P(4)^1 = \left(\frac{1}{5}\right)^5 = \frac{1}{3125}.$$

2. log-likelihood:
O log-likelihood é dado por:

$$l(U|D) = \log L(U|D) = 5 \log \left(\frac{1}{5}\right) = -5 \log(5).$$

Questão 4

Assuma que você tem uma moeda viciada tal que com probabilidade p você obtém caras (H) e $(1 - p)$ coroas (T). Você joga a moeda N vezes e obtém N_H caras (e $N - N_H$ coroas, é claro).

1. Obtenha a função de verossimilhança $L(\theta|D) = p(D|\theta)$ onde θ é o vetor de parâmetros. Qual é o valor de $p(D|\theta)$ se $D = \{HHTHTTHTTT\}$ e $p = 0.2$? E se $p = 0.6$?

2. Para D dado no item acima, encontre p que otimiza o log-likelihood. De maneira geral, encontre p como uma função de N e N_H para qualquer conjunto D dado.

Resposta:

Questão 4
1. Verossimilhança:
A função de verossimilhança para N_H caras em N lançamentos é dada por:

$$L(p|D) = p^{N_H}(1-p)^{N-N_H}.$$

Para $D = \{HHT, HT, T, HT, T, T\}$ com $N = 6$ e $N_H = 3$:

$$L(p|D) = p^3(1-p)^3.$$

Calculando:
- Para $p = 0.2$:

$$L(0.2|D) = (0.2)^3(0.8)^3 \approx 0.004096.$$

- Para $p = 0.6$:

$$L(0.6|D) = (0.6)^3(0.4)^3 \approx 0.013824.$$

2. Otimização da log-verossimilhança
A log-verossimilhança é:

$$\log L(p|D) = N_H \log(p) + (N - N_H)\log(1-p).$$

A derivada em relação a p é:

$$\frac{d}{dp}\log L(p|D) = \frac{N_H}{p} - \frac{N - N_H}{1-p}.$$

Portanto, o valor que maximiza a log-verossimilhança é $p = \frac{N_H}{N}$.

Questão 5

Suponha agora que suas amostras são obtidas ou de uma distribuição discreta $U(1,5)$ ou a partir de um dado (seis faces), sendo que todas as amostras são obtidas da mesma distribuição. Suponha que a probabilidade das amostras serem obtidas do dado é igual a p. Considere o conjunto de dados $\{2, 2, 4, 3, 2\}$, e seja $p = 0.7$.

1. Qual a likelihood das amostras serem retiradas a partir: (a) do dado se seis faces, ou (b) de uma $U(1,5)$ discreta?

Resposta:

(a) **Verossimilhança das amostras para o dado de 6 faces $U(1,6)$:**

$$L(U(1,6)|\{2,2,4,3,2\}) = P(dados|U(1,6)) = \left(\frac{1}{6}\right)^5 = \frac{1}{7776}.$$

(b) **Verossimilhança das amostras para $U(1,5)$:**

$$L(U(1,5)|\{2,2,4,3,2\}) = P(dados|U(1,5)) = \left(\frac{1}{5}\right)^5 = \frac{1}{3125}.$$

2. Qual a distribuição posterior?

Resposta:

Usando o Teorema de Bayes, temos:

$$P(U(1,\theta)|\text{dados}) = \frac{P(\text{dados}|U(1,\theta))P(U(1,\theta))}{P(\text{dados})}$$

Onde θ é o limite superior da distribuição uniforme e $P(\text{dados})$ é a probabilidade total de observar os dados, dada por:

$$P(\text{dados}) = P(\text{dados}|U(1,5))P(U(1,5)) + P(\text{dados}|U(1,6))P(U(1,6))$$

Como $P(U(1,6)) = 0.7$, substituímos as verossimilhanças:

$$P(U(1,6)|\text{dados}) = \frac{P(\text{dados}|U(1,6))P(U(1,6))}{P(\text{dados})} = \frac{1/7776 \cdot 0.7}{1/7776 \cdot 0.7 + 1/3125 \cdot 0.3} \approx 0.4839.$$

E para $U(1,5)$:

$$P(U(1,5)|\text{dados}) = 1 - P(U(1,6)|\text{dados}) \approx 0.5161.$$

(a) $P(U(1,6)|\text{dados}) = 0.4839$

(b) $P(U(1,5)|\text{dados}) = 0.5161$

3. Uma vez que o dataset acima foi observado, qual a probabilidade de se retirar o número 5?

Resposta:

A probabilidade de retirar o número 5, após observar o conjunto de dados {2, 2, 4, 3, 2}, é uma combinação ponderada das probabilidades de retirar o número 5 tanto de $U(1,5)$ quanto de $U(1,6)$, usando as probabilidades a posteriori calculadas no item anterior.
A probabilidade de retirar o número 5 de $U(1,5)$ é $\frac{1}{5}$ e de $U(1,6)$ é $\frac{1}{6}$. Assim, a probabilidade total é:

$$P(5|\text{dados}) = P(5|U(1,5)) \cdot P(U(1,5)|\text{dados}) + P(5|U(1,6)) \cdot P(U(1,6)|\text{dados})$$

Substituindo os valores:

$$P(5|\text{dados}) = \frac{1}{5} \cdot 0.5161 + \frac{1}{6} \cdot 0.4839 = 0.1032 + 0.0806 = 0.1838.$$

Portanto, a probabilidade de retirar o número 5 é aproximadamente 0.1838.

4. Uma vez que o dataset acima foi observado, qual a probabilidade de se retirar o número 6?

Resposta:

A probabilidade de retirar o número 6, após observar o conjunto de dados {2, 2, 4, 3, 2}, é uma combinação ponderada das probabilidades de retirar o número 6 tanto de $U(1,5)$ quanto de $U(1,6)$, usando as probabilidades a posteriori calculadas no item anterior. Entretanto, a probabilidade de retirar o número 6 de $U(1,5)$ é zero, e de $U(1,6)$ é $\frac{1}{6}$. Assim, a probabilidade total é:

$$P(6|\text{dados}) = P(6|U(1,5)) \cdot P(U(1,5)|\text{dados}) + P(6|U(1,6)) \cdot P(U(1,6)|\text{dados})$$

Substituindo os valores:

$$P(6|\text{dados}) = 0 \cdot 0.5161 + \frac{1}{6} \cdot 0.4839 = 0 + 0.0806 = 0.0806.$$

Portanto, a probabilidade de retirar o número 6 é aproximadamente 0.0806.

5. Qual o MLE?

Resposta:

As amostras $\{2, 2, 4, 3, 2\}$ vêm de uma distribuição uniforme $U(1, \theta)$. O maior valor observado é 4, então $\theta \geq 4$.
A função de verossimilhança é:

$$L(\theta) = \prod_{i=1}^{n} P(x_i | U(1, \theta)) = \left(\frac{1}{\theta}\right)^n$$

A log-verossimilhança é:

$$\log L(\theta) = -n \cdot \log(\theta)$$

Para maximizar a log-verossimilhança, derivamos em relação a θ e igualamos a zero:

$$\frac{d}{d\theta} \log L(\theta) = -\frac{n}{\theta} = 0$$

Como $n = 5$, e $\theta \geq 4$, essa igualdade nunca será válida. Portanto, para maximizar a verossimilhança, θ deve ser o menor valor possível, ou seja, $\theta = 4$.
Portanto, o MLE é $\theta = 4$.

6. Qual o MAP?

Resposta:

Para calcular o MAP, utilizamos o Teorema de Bayes:

$$P(\theta | D) \propto P(D | \theta) \cdot P(\theta)$$

Onde:

- $P(\theta | D)$ é a posteriori sobre θ.
- $P(D | \theta) = \left(\frac{1}{\theta}\right)^n$ é a verossimilhança.
- $P(\theta)$ é a priori sobre θ.

Assumindo uma priori uniforme $P(\theta)$ sobre $[4, \infty)$, temos:

$$P(\theta | D) \propto \left(\frac{1}{\theta}\right)^n$$

O MAP ocorre onde a posteriori é maximizada. Como a verossimilhança diminui com o aumento de θ, o valor que maximiza $P(\theta | D)$ é o menor valor permitido:

$$\hat{\theta}_{MAP} = 4$$

7. Caso $p = 0.5$, quais das respostas acima mudam de valor? Explique.

Resposta:

(a) **Verossimilhança:**

- A verossimilhança das amostras para o dado de 6 faces $U(1, 6)$ permanece a mesma, pois a distribuição das amostras não muda.
- A verossimilhança das amostras para $U(1, 5)$ também permanece a mesma.

(b) **Distribuição posterior:**

- A distribuição posterior muda, pois a probabilidade de que as amostras venham do dado de 6 faces é agora de 50%.

(c) **Probabilidade de retirar o número 5:**

- A probabilidade de retirar o número 5 muda, pois a probabilidade a posteriori de que as amostras venham do dado de 6 faces é agora de 50%.

(d) **Probabilidade de retirar o número 6:**

- A probabilidade de retirar o número 6 muda, pois a probabilidade a posteriori de que as amostras venham do dado de 6 faces é agora de 50%.

(e) **MLE:**

- O MLE não muda, pois a verossimilhança é maximizada quando $\theta = 4$.

(f) **MAP:**

- O MAP não muda, pois foi assumida que a distribuição a priori é uniforme. Logo a posteriori é maximizada quando $\theta = 4$.

Questão 6

Suponha agora que suas amostras são obtidas ou de uma distribuição discreta $U(1, 5)$ ou a partir de um dado (seis faces) com probabilidade $(1 - p)$ e p, respectivamente. Entretanto, nesta questão, a sequência pode conter amostras de ambas distribuições (mistura de distribuições). Considere o mesmo conjunto de dados $\{2, 2, 4, 3, 2\}$, e seja $p = 0.7$.

1. Qual a likelihood de observar essas amostras?

Resposta:

Para analisar o conjunto de dados $\{2, 2, 4, 3, 2\}$, que pode ter sido gerado a partir de uma mistura das distribuições $U(1,5)$ e um dado (seis faces), com $p = 0.7$ para o dado e $1 - p = 0.3$ para a distribuição uniforme, seguimos os seguintes passos:

- **Probabilidades dos Valores**

  As probabilidades de cada valor no conjunto de dados são calculadas da seguinte forma:

  Para $x = 2$:
  $$P(2) = 0.3 \cdot \frac{1}{5} + 0.7 \cdot \frac{1}{6} \approx 0.17667$$
  Para $x = 4$:
  $$P(4) = 0.3 \cdot \frac{1}{5} + 0.7 \cdot \frac{1}{6} \approx 0.17667$$
  Para $x = 3$:
  $$P(3) = 0.3 \cdot \frac{1}{5} + 0.7 \cdot \frac{1}{6} \approx 0.17667$$

- **Cálculo da Likelihood**

  A likelihood de observar as amostras $\{2, 2, 4, 3, 2\}$ é dada por:
  $$L = P(2)^3 \cdot P(4)^1 \cdot P(3)^1$$
  Substituindo os valores:
  $$L = (0.17667)^3 \cdot (0.17667)^1 \cdot (0.17667)^1 = (0.17667)^5 \approx 0.000179$$

2. Uma vez que o dataset acima foi observado, qual a probabilidade de se retirar o número 5?

Resposta:

Para calcular a probabilidade de se retirar o número 5 do conjunto de dados {2, 2, 4, 3, 2} com base nas distribuições misturadas $U(1,5)$ e $U(1,6)$, precisamos considerar como cada distribuição contribui para a probabilidade de obter o número 5.

$$P(5 \mid U(1,5)) = \frac{1}{5} = 0.2$$

$$P(5 \mid U(1,6)) = \frac{1}{6} \approx 0.16667$$

A probabilidade total de observar o número 5 pode ser calculada usando a fórmula de mistura:

$$P(5) = p \cdot P(5 \mid U(1,6)) + (1-p) \cdot P(5 \mid U(1,5))$$

Onde $p = 0.7$ (para $U(1,6)$) e $1 - p = 0.3$ (para a distribuição uniforme $U(1,5)$). Substituindo os valores:

$$P(5) = 0.7 \cdot \frac{1}{6} + 0.3 \cdot \frac{1}{5} = 0.11667 + 0.06 = 0.17667$$

3. Uma vez que o dataset acima foi observado, qual a probabilidade de se retirar o número 6?

Resposta:

Seguimos os mesmos passos do item anterior, levando em consideração que na distribuição uniforme $U(1,5)$, o número 6 não está presente, portanto:

$$P(6 \mid U(1,5)) = 0$$

$$P(6 \mid U(1,6)) = \frac{1}{6} \approx 0.16667$$

A probabilidade total de observar o número 6 pode ser calculada usando a fórmula de mistura:

$$P(6) = p \cdot P(6 \mid U(1,6)) + (1-p) \cdot P(6 \mid U(1,5))$$

Substituindo os valores:

$$P(6) = 0.7 \cdot \frac{1}{6} + 0.3 \cdot 0 = 0.11667 + 0 \approx 0.11667$$

4. Qual a probabilidade de **todas** as amostras serem retiradas a partir: (a) do dado se seis faces, ou (b) de uma $U(1,5)$ discreta?

Resposta:

Para calcular a probabilidade de que todas as amostras do conjunto $\{2, 2, 4, 3, 2\}$ sejam retiradas a partir do dado com seis faces ou da distribuição uniforme $U(1,5)$, precisamos analisar cada caso separadamente.

(a) **Probabilidade de todas as amostras serem retiradas do dado** $U(1,6)$:

A probabilidade de observar cada número x na distribuição $U(1,6)$ é:

$$P(x \mid U(1,6)) = \frac{1}{6}$$

Para as amostras $\{2, 2, 4, 3, 2\}$, a probabilidade total é:

$$P(D \mid U(1,6)) = P(2 \mid U(1,6))^3 \cdot P(4 \mid U(1,6)) \cdot P(3 \mid U(1,6)) = \left(\frac{1}{6}\right)^5$$

$$P(D \mid U(1,6)) \approx 0.0001286$$

(b) **Probabilidade de todas as amostras serem retiradas da distribuição** $U(1,5)$:

A probabilidade de observar cada número x na distribuição $U(1,5)$ é:

$$P(x \mid U(1,6)) = \frac{1}{5}$$

Para as amostras $\{2, 2, 4, 3, 2\}$, a probabilidade total é:

$$P(D \mid U(1,5)) = P(2 \mid U(1,5))^3 \cdot P(4 \mid U(1,5)) \cdot P(3 \mid U(1,5)) = \left(\frac{1}{5}\right)^5$$

$$P(D \mid U(1,5)) \approx 0.00032$$

5. Calcule a função de likelihood para as amostras em função de p, o log likelihood e obtenha o valor de p que melhor explica o conjunto de dados. Comente a sua resposta.

Resposta:

Para calcular a função de likelihood $L(p)$ e o log likelihood $\ell(p)$ para as amostras $\{2$, 2, 4, 3, $2\}$, utilizamos a seguinte abordagem:
A função de likelihood é dada por:

$$L(p) = P(\{2, 2, 4, 3, 2\} \mid p) = \prod_{i=1}^{n} (p \cdot P(x_i \mid U(1, 6)) + (1 - p) \cdot P(x_i \mid U(1, 5)))$$

Para as amostras em questão:

$$L(p) = \left(p \cdot \frac{1}{6} + (1 - p) \cdot \frac{1}{5}\right)^5$$

O log likelihood é:

$$\ell(p) = \log(L(p)) = 5 \log \left(p \cdot \frac{1}{6} + (1 - p) \cdot \frac{1}{5}\right)$$

Derivando $\ell(p)$ em relação a p e igualando a zero, obtemos o valor de p que maximiza o log likelihood:

$$\frac{d}{dp}\ell(p) = \frac{5}{p \cdot 6 + (1 - p) \cdot 5} \left(\frac{1}{6} - \frac{1}{5}\right) = 0$$

Como podemos observar, a igualdade acima não é válida para nenhum valor de p. Portanto, não é possível determinar o valor de p que melhor explica o conjunto de dados.

6. Repita o item anterior, supondo que o conjunto de dados tem cardinalidade 20 e apenas uma única amostra tenha valor igual a 6.

Resposta:

A função de likelihood é dada por:

$$L(p) = P(D \mid p) = \prod_{i=1}^{n} (p \cdot P(x_i \mid U(1, 6)) + (1 - p) \cdot P(x_i \mid U(1, 5)))$$

Para as amostras em questão:

$$L(p) = \left(p \cdot \frac{1}{6} + (1 - p) \cdot \frac{1}{5}\right)^{19} \cdot \left(p \cdot \frac{1}{6} + (1 - p) \cdot 0\right)$$