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

Departamento de Engenharia Elétrica
COE782 - Introdução ao Aprendizado de Máquina
Prof. Dr. Markus Vinícius Santos Lima

Lista 2 de exercícios

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

2 de junho de 2024

Exercícios do Bishop

1. **Exercício 2.1**

   Verifique que a distribuição de Bernoulli $Bern(x|\mu) = \mu^x(1-\mu)^{1-x}$ satisfaz as seguintes propriedades:

   $$\sum_{x=0}^{1} p(x|\mu) = 1$$
   $$\mathbb{E}[x] = \mu$$
   $$var[x] = \mu(1-\mu).$$

   Mostre que a entropia $H[x]$ de uma variável binária x com distribuição Bernoulli é dada por

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

   Solução:

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

   $\mathbb{E}[x] = \sum_{x=0}^{1} p(x|\mu)x = \sum_{x=0}^{1} \mu^x(1-\mu)^{1-x}x = \mu^1(1-\mu)^{(1-1)}1 + \mu^0(1-\mu)^{(1-0)}0 = \boxed{\mu}$

   $var[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sum_{x=0}^{1} \mu^x(1-\mu)^{1-x}x^2 - \mu^2 = \mu - \mu^2 = \boxed{\mu(1-\mu)}$

   $H[x] = -\sum_x p(x)\ln p(x) = -\sum_x \mu^x(1-\mu)^{1-x} \ln \mu^x(1-\mu)^{1-x}$

   $H[x] = -\mu^1(1-\mu)^{1-1}\ln \mu^1(1-\mu)^{1-1} - \mu^0(1-\mu)^{1-0}\ln \mu^0(1-\mu)^{1-0}$

   $H[x] = \boxed{-\mu \ln \mu - (1-\mu)\ln(1-\mu)}$

2. **Exercício 2.2**

   A forma da distribuição de Bernoulli dada por $Bern(x|\mu) = \mu^x(1-\mu)^{1-x}$ não é simétrica entre os dois valores de x. Em algumas situações, será mais conveniente usar uma formulação equivalente para a qual ($x \in \{-1, 1\}$), caso em que a distribuição pode ser escrita

   $$p(x|\mu) = \left(\frac{1-\mu}{2}\right)^{(1-x)/2} \left(\frac{1+\mu}{2}\right)^{(1+x)/2}$$

   onde $\mu \in [-1, 1]$. Mostre que a distribuiição de $p(x|\mu)$ é normalizada e avalie sua média, variância e entropia.

   Solução:

   $\sum_x p(x|\mu) = \left(\frac{1-\mu}{2}\right)^{(1-(-1))/2} \left(\frac{1+\mu}{2}\right)^{(1+(-1))/2} + \left(\frac{1-\mu}{2}\right)^{(1-(1))/2} \left(\frac{1+\mu}{2}\right)^{(1+(1))/2}$

   $\sum_x p(x|\mu) = \left(\frac{1-\mu}{2}\right) + \left(\frac{1+\mu}{2}\right) = \boxed{1}$

   $\mathbb{E}[x] = \sum_{x=0}^{1} p(x|\mu)x = \sum_{x=-1}^{1} \left(\frac{1-\mu}{2}\right)^{(1-x)/2} \left(\frac{1+\mu}{2}\right)^{(1+x)/2} x$

$$\mathbb{E}[x] = \left(\frac{1-\mu}{2}\right)^{(1-(-1))/2}\left(\frac{1+\mu}{2}\right)^{(1+(-1))/2}(-1) + \left(\frac{1-\mu}{2}\right)^{(1-(1))/2}\left(\frac{1+\mu}{2}\right)^{(1+(1))/2}(1)$$

$$\mathbb{E}[x] = \left(\frac{1-\mu}{2}\right)(-1) + \left(\frac{1+\mu}{2}\right)(1)$$

$$\mathbb{E}[x] = -\frac{1-\mu}{2} + \frac{1+\mu}{2} = \boxed{\mu}$$

$$\mathbb{E}[x^2] = \textstyle\sum_{x=0}^{1} p(x|\mu)x^2 = \sum_{x=-1}^{1} \left(\frac{1-\mu}{2}\right)^{(1-x)/2}\left(\frac{1+\mu}{2}\right)^{(1+x)/2} x^2$$

$$\mathbb{E}[x^2] = \left(\frac{1-\mu}{2}\right)^{(1-(-1))/2}\left(\frac{1+\mu}{2}\right)^{(1+(-1))/2}(-1)^2 + \left(\frac{1-\mu}{2}\right)^{(1-(1))/2}\left(\frac{1+\mu}{2}\right)^{(1+(1))/2}(1)^2$$

$$\mathbb{E}[x^2] = \left(\frac{1-\mu}{2}\right)(1) + \left(\frac{1+\mu}{2}\right)(1)$$

$$\mathbb{E}[x^2] = \frac{1-\mu}{2} + \frac{1+\mu}{2} = 1$$

$$var[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \boxed{1-\mu^2}$$

$$H[x] = -\textstyle\sum_x p(x)\ln p(x) = \boxed{-\left(\frac{1-\mu}{2}\right)\ln\left(\frac{1-\mu}{2}\right) - \left(\frac{1+\mu}{2}\right)\ln\left(\frac{1+\mu}{2}\right)}$$

3. **Exercício 2.8**

   Considere duas variáveis x e y com distribuição conjunta $p(x, y)$. Prove o seguinte resultado (resultado bem útil conhecido como Regra da Torre)

   $$\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x|y]].$$

   Aqui, $\mathbb{E}_x[x|y]$ denota o valor esperado de x sob a distribuição condicional $p(x|y)$.

   Solução:

   $\mathbb{E}[x] = \int_y \int_x xp(x, y)dydx$

   Mas $p(x, y) = p(x|y)p(y)$ (regra do produto). Então,

   $\mathbb{E}[x] = \int_y \int_x xp(x|y)p(y)dydx = \int_y p(y)\left[\int_x xp(x|y)dx\right]dy$

   $\mathbb{E}[x] = \int_y p(y)\mathbb{E}_x[x|y]dy$

   $\mathbb{E}[x] = \boxed{\mathbb{E}_y[\mathbb{E}_x[x|y]]}$

4. **Exercício 2.12**

   A distribuição uniforme de uma variável contínua x é definida por

   $$U(x|a, b) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

   Verifique se esta distribuição está normalizada e encontre expressões para sua média e variância.

   Solução:

$$\int_a^b U(x|a,b)dx = \int_a^b \frac{1}{b-a}dx = \frac{1}{b-a}[x]_a^b = \frac{1}{b-a}(b-a) = \boxed{1}$$

$$\mathbb{E}[x] = \int_a^b \frac{1}{b-a}xdx = \frac{1}{b-a}\left[\frac{x}{2}\right]_a^b = \frac{1}{b-a}\left(\frac{b^2-a^2}{2}\right) = \frac{(b-a)(b+a)}{2(b-a)} = \boxed{\frac{a+b}{2}}$$

$$\mathbb{E}[x^2] = \int_a^b \frac{1}{b-a}x^2dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{b-a}\left(\frac{b^3-a^3}{2}\right) = \frac{(b-a)(a^2+ab+b^2)}{3(b-a)} = \frac{(a^2+ab+b^2)}{3}$$

$$var[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \frac{(a^2+ab+b^2)}{3} - \frac{(a+b)^2}{4} = \frac{4(a^2+ab+b^2) - 3(a^2+2ab+b^2)}{12}$$

$$var[x] = \frac{a^2-2ab+b^2}{12} = \boxed{\frac{(b-a)^2}{12}}$$

5. **Exercício 2.13**

   Avalie a divergência de Kullback-Leibler

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

   entre duas distribuições Gaussianas $p(x) = \mathcal{N}(x|\mu,\Sigma)$ e $q(x) = \mathcal{N}(x|m,L)$ para o caso geral e para os seguintes casos:

   (a) ambas as pdfs têm mesma média e matriz de covariância (isto é, p(x) = q(x); caso em que já sabemos quanto a divergência KL deve resultar);

   (b) ambas as pds têm a mesma média, isto é, $m = \mu$..

   Solução geral:

$$p(x) = \mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

$$q(x) = \mathcal{N}(x|m,L) = \frac{1}{(2\pi)^{N/2}|L|^{1/2}}\exp\left(-\frac{1}{2}(x-m)^T L^{-1}(x-m)\right)$$

$$KL(p||q) = -\int p(x)\ln\left\{\tfrac{q(x)}{p(x)}\right\}dx$$

$$= -\int p(x)\ln\left\{\tfrac{|\Sigma|^{1/2}}{|L|^{1/2}}\exp\left(-\tfrac{1}{2}(x-m)^T L^{-1}(x-m) + \tfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)\right\}dx$$

$$= -\int p(x)\ln\left\{\tfrac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\}dx - \int p(x)\ln\left\{\exp\left(-\tfrac{1}{2}(x-m)^T L^{-1}(x-m) + \tfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)\right\}dx$$

$$= -\ln\left\{\tfrac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} - \int p(x)\left(-\tfrac{1}{2}(x-m)^T L^{-1}(x-m) + \tfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dx$$

$$= -\ln\left\{\tfrac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} - \int p(x)\left(-\tfrac{1}{2}(x-m)^T L^{-1}(x-m)\right)dx - \int p(x)\left(\tfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dx$$

$$KL(p||q) = -\ln\left\{\tfrac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} - \mathbb{E}\left[-\tfrac{1}{2}(x-m)^T L^{-1}(x-m)\right] - \mathbb{E}\left[\tfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

$$KL(p||q) = -\ln\left\{\tfrac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} + \tfrac{1}{2}\mathbb{E}\left[(x-m)^T L^{-1}(x-m)\right] - \tfrac{1}{2}\mathbb{E}\left[(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

$KL(p||q) = -\ln\left\{\frac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} + \frac{1}{2}tr\left(L^{-1}\mathbb{E}\left[(x-m)^T(x-m)\right]\right) - \frac{1}{2}tr\left(\Sigma^{-1}\mathbb{E}\left[(x-\mu)^T(x-\mu)\right]\right)$

$KL(p||q) = -\ln\left\{\frac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} + \frac{1}{2}tr\left(L^{-1}\mathbb{E}\left[(x-m)^T(x-m)\right]\right) - \frac{1}{2}tr\left(\Sigma^{-1}\mathbb{E}\left[(x-\mu)^T(x-\mu)\right]\right)$

Como μ é a média de $p(x)$, temos

$KL(p||q) = -\ln\left\{\frac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} + \frac{1}{2}tr\left(L^{-1}\mathbb{E}\left[(x-m)^T(x-m)\right]\right) - \frac{1}{2}tr\left(\Sigma^{-1}\Sigma\right)$

$KL(p||q) = -\ln\left\{\frac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} - \frac{1}{2}tr\left(I\right) + \frac{1}{2}tr\left(L^{-1}\mathbb{E}\left[(x-m)^T(x-m)\right]\right)$

$KL(p||q) = \underline{-\ln\left\{\frac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} - \frac{D}{2} + \frac{1}{2}tr\left(L^{-1}\mathbb{E}\left[(x-m)^T(x-m)\right]\right)}$

Solução para $\Sigma = L$ e $\mu = m$:

$KL(p||q) = -\ln\left\{\frac{|\Sigma|^{1/2}}{|\Sigma|^{1/2}}\right\} - \frac{D}{2} + \frac{1}{2}tr\left(\Sigma^{-1}\mathbb{E}\left[(x-\mu)^T(x-\mu)\right]\right)$

$KL(p||q) = -\ln\{1\} - \frac{D}{2} + \frac{1}{2}tr\left(\Sigma^{-1}\Sigma\right)$

$KL(p||q) = -0 - \frac{D}{2} + \frac{1}{2}tr\left(I\right)$

$KL(p||q) = -\frac{D}{2} + \frac{D}{2}$

$KL(p||q) = \underline{0}$

Solução para apenas $\mu = m$:

$KL(p||q) = -\ln\left\{\frac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} - \frac{D}{2} + \frac{1}{2}tr\left(L^{-1}\mathbb{E}\left[(x-\mu)^T(x-\mu)\right]\right)$

$KL(p||q) = -\ln\left\{\frac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} - \frac{D}{2} + \frac{1}{2}tr\left(L^{-1}L\right)$

$KL(p||q) = -\ln\left\{\frac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} - \frac{D}{2} + \frac{1}{2}tr\left(I\right)$

$KL(p||q) = -\ln\left\{\frac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\} - \frac{D}{2} + \frac{D}{2}$

$KL(p||q) = \underline{-\ln\left\{\frac{|\Sigma|^{1/2}}{|L|^{1/2}}\right\}}$

6. **Exercício 2.15**

   Mostre que a entropia da Gaussiana multivariada $\mathcal{N}(x|\mu,\Sigma)$ é dada por

   $$H[x] = \frac{1}{2}\ln|\Sigma| + \frac{D}{2}(1 + \ln(2\pi))$$

   onde D é a dimensionalidade de x.

   Solução:

   $H[x] = -\int p(x)\ln p(x)dx$ e $p(x) = \mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$

   Simplificando primeiro o ln

   $\ln p(x) = \ln\left[\frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)\right]$

   $\ln p(x) = -\dfrac{D}{2}\ln[2\pi] - \dfrac{1}{2}\ln|\Sigma| - \dfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$

Substituindo na fórmula da entropia

$H[x] = -\int p(x)\left(-\frac{D}{2}\ln[2\pi] - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dx$

$H[x] = -\int p(x)\left(-\frac{D}{2}\ln[2\pi]\right)dx + \int p(x)\left(-\frac{1}{2}\ln|\Sigma|\right)dx + \int p(x)\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dx$

$H[x] = -\left(-\frac{D}{2}\ln[2\pi]\right) - \left(-\frac{1}{2}\ln|\Sigma|\right) - \int p(x)\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dx$

$H[x] = \frac{D}{2}\ln[2\pi] + \frac{1}{2}\ln|\Sigma| - \int p(x)\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)dx$

$H[x] = \frac{D}{2}\ln[2\pi] + \frac{1}{2}\ln|\Sigma| + \frac{1}{2}\mathbb{E}\left[(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$

$H[x] = \frac{D}{2}\ln[2\pi] + \frac{1}{2}\ln|\Sigma| + \frac{1}{2}tr\left(\Sigma^{-1}\mathbb{E}\left[(x-\mu)^T(x-\mu)\right]\right)$

$H[x] = \frac{D}{2}\ln[2\pi] + \frac{1}{2}\ln|\Sigma| + \frac{1}{2}tr\left(\Sigma^{-1}\Sigma\right) = \frac{D}{2}\ln[2\pi] + \frac{1}{2}\ln|\Sigma| + \frac{1}{2}tr\left(I\right)$

$H[x] = \frac{D}{2}\ln[2\pi] + \frac{1}{2}\ln|\Sigma| + \frac{D}{2} = \underline{\frac{1}{2}\ln|\Sigma| + \frac{D}{2}\left(1+\ln[2\pi]\right)}\Big|$

7. **Exercício 2.20**

Uma matriz definida positiva Σ pode ser definida como aquela para a qual o forma quadrática

$$a^T\Sigma a$$

é positiva para qualquer valor real do vetor a. Mostre que uma condição necessária e suficiente para Σ ser definida positiva é que todos os autovalores λ_i de Σ, definidos por $\Sigma u_i = \lambda_i u_i$, sejam positivos.

Solução:

Como u_1,...,u_D constituem uma base para o R^D, podemos escrever

$$a = a_1u_1 + a_2u_2 + ... + a_Du_D$$

onde a_1,...,a_D são coeficientes obtidos projetando-se a em u_1,...,u_D.

Então,

$$a^T\Sigma a = (a_1u_1^T + ... + a_Du_D^T)\Sigma(a_1u_1 + ... + a_Du_D)$$

Aplicando a definição $\Sigma u_i = \lambda_i u_i$, temos

$$a^T\Sigma a = (a_1u_1^T + ... + a_Du_D^T)(a_1\lambda_1u_1 + ... + a_D\lambda_Du_D)$$

Como $u_i^Tu_j = 1$ apenas para $i = j$, sendo 0 caso contrário, temos

$$a^T \Sigma a = (a_1^2 \lambda_1 + ... + a_D^2 \lambda_D)$$

Se os autovalores λ_j são sempre positivos, $a_j^2 \lambda_j \geq 0$ para todo e qualquer j, pois a_j será sempre um número real por ser o coeficiente de uma projeção.

Logo, se todos os autovalores λ_j de Σ são positivos, Σ é positiva definida.

Exercícios do Bishop

E1 **Inferência Bayesiana sequencial**

Motivado pela Figura 2.3 do livro, reproduza o experimento da jogada de moeda considerando que foram realizadas 5 jogadas e que a probabilidade de se obter cara é dada por '$\mu = 0,7$'. Plote a distribuição a priori e todas as 5 distribuições a posteriori geradas ao longo do processo iterativo. Considere que a distribuição a priori é uma Beta com parâmetros 'a' e 'b' escolhidos da seguinte forma:

1º caso: $a = b = 1$

2º caso: $a = b = 2$

Compare os resultados obtidos nos 2 casos.

OBS: Para uma comparação justa entre os 2 casos, primeiro gere os 5 dados (saídas do experimento da moeda, amostrados da Bernoulli definida no enunciado) e depois aplique o aprendizado sequencial para os 2 casos (i.e., para ambas as prioris) usando exatamente os mesmos dados gerados.

Figura 2.3 do Bishop

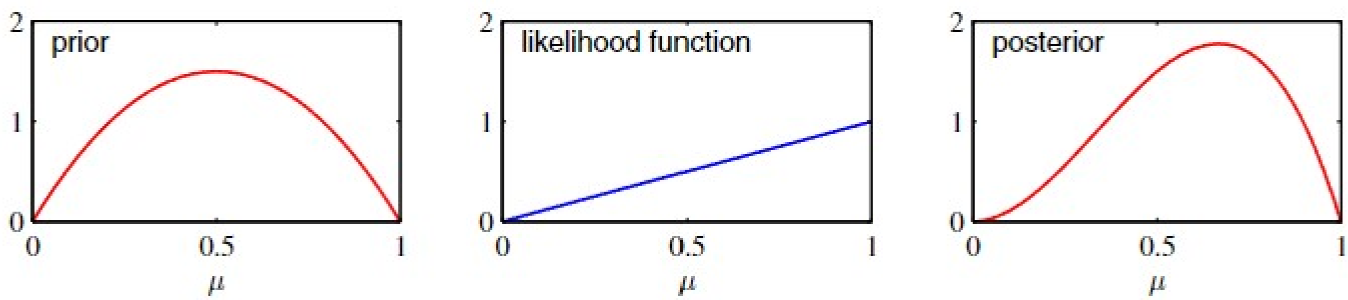


Figure 2.3 Illustration of one step of sequential Bayesian inference. The prior is given by a beta distribution with parameters $a = 2$, $b = 2$, and the likelihood function, given by (2.9) with $N = m = 1$, corresponds to a single observation of $x = 1$, so that the posterior is given by a beta distribution with parameters $a = 3$, $b = 2$.

Solução:

E2 **Verificação experimental do Teorema Central do Limite**

Considere a média de N variáveis aleatórias iid. Plote o histograma dessa média considerando que as N variáveis aleatórias têm a seguinte pdf:

1o caso: Uniforme(0,1) - uniforme no intervalo 0 a 1;

2o caso: Bernoulli - escolha o valor do parâmetro como quiser;

Note que, para N suficientemente grande, a distribuição da média converge para uma Gaussiana.

OBS: Usei média ao invés de soma para facilitar a geração do histograma (o eixo horizontal vai ficar fixo, facilitando a comparação para diferentes valores de N, igual na Figura 2.6 do livro).

Figura 2.6 do Bishop

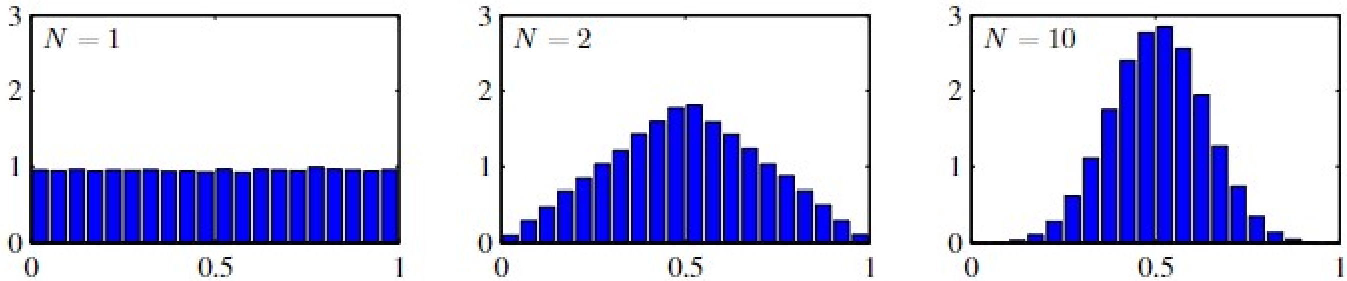


Figure 2.6 Histogram plots of the mean of N uniformly distributed numbers for various values of N. We observe that as N increases, the distribution tends towards a Gaussian.

Solução:

E3 **Verificação experimental da Law of Large Numbers - LLN**

Considere N variáveis aleatórias independentes geradas a partir de uma distribuição normal padrão (isto é, Gaussiana de média 0 e variância 1). Compute o estimador média amostral. Repita o experimento diversas vezes e plote o histograma desse estimador para um dado valor de N. Mostre que, conforme N aumenta, o histograma desse estimador fica cada vez mais estreito em torno do valor correto (i.e., variância vai diminuindo), que é zero.

Solução:

E4 **Estimação de pdf**

Tente replicar os resultados exibidos nas figuras 2.24 e 2.25 do livro. Para isso, gere uma amostra de N=50 dados cuja distribuição é dada pela curva em verde (corresponde a uma mistura de 2 gaussianas - veja equação

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k.\Sigma_k)$$

e escolha os parâmetros dessa distribuição da forma que quiser). Estime a pdf do modelo gerador dos dados utilizando 2 métodos: histograma e kernel Gaussiano. Para o parâmetro h, utilize os mesmos valores das figuras.

Figura 2.24 do Bishop

Figure 2.24 An illustration of the histogram approach to density estimation, in which a data set of 50 data points is generated from the distribution shown by the green curve. Histogram density estimates, based on (2.241), with a common bin width Δ are shown for various values of Δ.

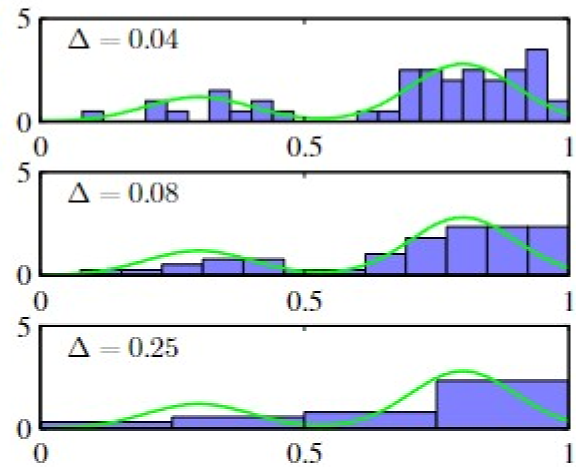
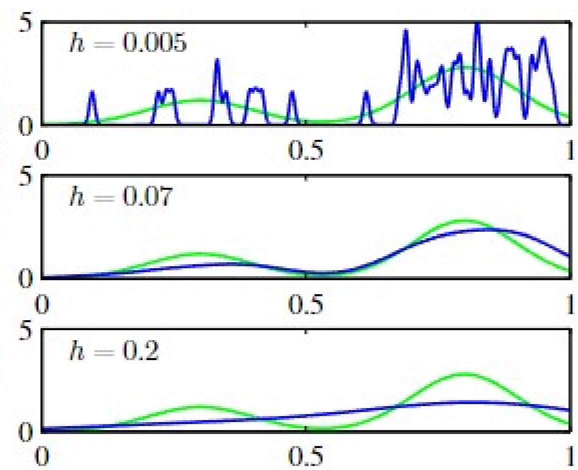


Figura 2.25 do Bishop

Figure 2.25 Illustration of the kernel density model (2.250) applied to the same data set used to demonstrate the histogram approach in Figure 2.24. We see that h acts as a smoothing parameter and that if it is set too small (top panel), the result is a very noisy density model, whereas if it is set too large (bottom panel), then the bimodal nature of the underlying distribution from which the data is generated (shown by the green curve) is washed out. The best density model is obtained for some intermediate value of h (middle panel).



Solução:

E5 **Classificador K-NN**

Considere 2 classes, C1 e C2, que correspondem aos seguintes modelos geradores:

C1: pdf Gaussiana de média -1 e variância 1

C2: pdf Gaussiana de média 1 e variância 1

Gere 10 observações de cada uma dessas classes e assuma que você sabe exatamente a classe de cada um dos 20 pontos gerados (10 pontos para cada classe). Em seguida, gere mais 2 observações de cada classe e assuma que você NÃO sabe de qual classe esses novos dados pertencem. Utilize a técnica de K-NN, considerando diferentes valores de K, para classificar os 4 novos dados.

OBS: Plote os resultados utilizando cores e símbolos para facilitar a interpretação. Por exemplo: Para os 20 pontos conhecidos, represente-os usando 'bolinhas' vermelhas para C1 e azuis para C2. Para os 4 pontos a serem classificados, mantenha o código de cores (para sabermos identificar qual era a classe correta) e use novos símbolos para identificar se a classificação foi correta (use um 'quadrado') ou se a classificação foi errada (neste caso, use um 'x').

Se for usar outros símbolos e cores, não tem problema. Só não esqueça de fazer uma legenda ou caption que me permita compreender a figura.

Solução: