DCC/ICEx/UFMG
Information Theory
2^{nd} semester of 2017

Prof. Mário S. Alvim
First Exam
2017/09/14

Student: Solução do Professor	Registration number:

Instructions:

i. This exam contains 6 page(s). If your exam is missing any page, please ask the instructor for another copy of the exam.

ii. You can use the back of the exam's sheets as a draft, but the final answer for each question must be provided within the designated boxes. **Only answers provided within these boxes will be considered for grading.**

iii. Your answers will be evaluated in their clarity and conciseness.

4

1. Preencha cada quadrado abaixo com V ou F de acordo com se a afirmativa correspondente é *verdadeira* ou *falsa*, respectivamente.

   Não é necessário justificar suas repostas, mas **cada resposta errada anulará uma correta.**

   Para as questões abaixo, considere os ensembles genéricos $X = (x, \mathcal{A}_X, \mathcal{P}_X)$ e $Y = (y, \mathcal{A}_Y, \mathcal{P}_Y)$.

   (a) $\boxed{F}$ Quando temos que decidir qual dentre duas hipóteses H_1 e H_2 é a melhor explicação para um mesmo conjunto de dados D, basta calcular a *likelihood ratio* $p(D|H_1)/p(D|H_2)$: se a razão for maior que 1, a hipótese H_1 deve ser escolhida. Se a razão for menor que 1, a hipótese H_2 deve ser escolhida. Se a razão for 1, qualquer hipótese pode ser escolhida.

   (b) $\boxed{V}$ Quanto maior o conteúdo de informação de Shannon $h(x)$ de um resultado x, mais surpreendente é este resultado.

   (c) $\boxed{V}$ Quanto maior a entropia $H(X)$ do ensemble, mais informação este ensemble, como um todo, carrega.

   (d) $\boxed{F}$ É possível que $H(X) < 0$, e neste caso dizemos que o ensemble carrega pouca informação.

   (e) $\boxed{F}$ É possível que $H(X) > |\mathcal{A}_X|$, e neste caso dizemos que o ensemble carrega muita informação.

   (f) $\boxed{V}$ Sempre que X e Y forem independentes, temos que $H(X,Y) = H(X) + H(Y)$.

   (g) $\boxed{V}$ Na compressão com perdas, todo arquivo pode ser comprimido em um arquivo de menor tamanho, mas o processo de descompressão pode conter erros.

   (h) $\boxed{F}$ É possível fazer compressão sem perdas de forma que todo arquivo seja comprimido em um arquivo de menor tamanho.

Comentários:

a) $\frac{p(D|H_1)}{p(D|H_2)}$ é apenas evidência. Para saber qual hipótese é mais provável, é preciso considerar $\frac{p(H_1)}{p(H_2)}$.

2. **(Tia Maude e as xícaras de chá.)** Sir Ronald Fisher foi um estatístico famoso que fez contribuições fundamentais para, dentre diversas outras áreas, a Teoria da Informação.

Reza a lenda que Sir Ronald Fisher tinha uma Tia Maude que gostava de seu chá das 5 com leite e açúcar. Tia Maude se gabava de ser capaz de diferenciar se sua xícara de chá tinha sido preparada adicionando-se o leite primeiro e depois o açúcar, ou adicionando-se primeiro o açúcar e depois o leite.

Para testar se Tia Maude falava a verdade ou não, Fisher preparou o seguinte experimento.

Ele vendou Tia Maude e preparou 8 xícaras de chá, sendo 5 delas preparadas da forma A (açúcar primeiro, depois leite), e 3 preparadas da forma B (leite primeiro, depois açúcar).

Fisher, então, aleatoriamente ordenou as xícaras para dar para Tia Maude provar, e anotou qual a resposta de Tia Maude, como na tabela abaixo.

xícara	1	2	3	4	5	6	7	8
tipo da xícara	A	A	B	A	B	A	A	B
avaliação de Tia Maude	A	B	A	B	A	B	A	B

Ou seja, a primeira xícara que Fisher deu para Tia Maude provar era to tipo A, e ela a avaliou como sendo do tipo A; a segunda xícara era do tipo A, e Tia Maude a avaliou como sendo do tipo B; etc.

O objetivo de Fisher é decidir qual de duas hipóteses é mais provável: ou Tia Maude fala a verdade, e que em 75% das vezes ela consegue identificar corretamente se o chá foi preparado da forma A ou da forma B; ou Tia Maude é um tanto quanto loroteira, e na verdade só consegue identificar corretamente em 50% das vezes a maneira como o chá foi preparado (o que seria esperado se ela "chutasse" ao acaso toda vez).

2

(a) Modele o problema de Fisher como um problema de comparação de modelos (ou teste de hipóteses), identificando quais os modelos a serem comparados, quais os dados disponíveis como evidência, e qual o critério matemático Fisher deve utilizar para decidir qual o melhor modelo.

As hipóteses são:

- H_0 = Tia Maude acerta com 50% de chance.
- H_1 = Tia Maude acerta com 75% de chance.

Os dados disponíveis como evidência são:

- D = A sequência de acertos e erros de Tia Maude.

O critério matemático a ser usado é computar

$$\frac{p(H_0 \mid D)}{p(H_1 \mid D)} = \frac{p(H_0)}{p(H_1)} \cdot \frac{p(D \mid H_0)}{p(D \mid H_1)},$$

onde $p(H_0)/p(H_1)$ é a razão das probabilidades a priori e $p(D \mid H_0)/p(D \mid H_1)$ é a razão de verossimilhança, e escolher H_1 como mais provável se $p(H_0 \mid D)/p(H_1 \mid D) > 1$, escolher H_2 como mais provável se $p(H_0 \mid D)/p(H_1 \mid D) < 1$, e ser indiferente a H_1 e H_2 se $p(H_0 \mid D)/p(H_1 \mid D) = 1$.

2 (b) A sequência de respostas dada por Tia Maude é evidência a favor ou contra ela estar falando a verdade? Justifique usando a *likelihood ratio*.

Computando a "likelihood ratio" (razão de verossimilhança):

$$\frac{p(D\mid H_0)}{p(D\mid H_1)} = \frac{1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2}{3/4 \cdot 1/4 \cdot 1/4 \cdot 1/4 \cdot 1/4 \cdot 1/4 \cdot 3/4 \cdot 3/4} = \frac{256}{27} \approx 9.48$$

Vemos que ela é maior que 1, e a evidência é a favor da hipótese H_0 (Tia Maude acerta com 50% de chance).

2 (c) Após o teste ser realizado, determine se Fisher pode concluir que sua Tia Maude fala a verdade nos seguintes casos.
Primeiro, se Tia Maude tem um histórico de ser uma pessoa honesta, e Fisher atribui uma probabilidade *a priori* de ela estar falando a verdade de 90%.
Segundo, se Fisher considera praticamente impossível que alguém tenha a habilidade que Tia Maude diz ter, e atribui uma probaiblide *a priori* de ela estar falando a verdade de apenas 10%.
Justifique sua calculando a probabilidade *a posteriori* de Tia Maude falar a verdade, para dois casos.

Primeiro caso: Tia Maude é honesta com 0.90 de chance:

$$\frac{p(H_0\mid D)}{p(H_1\mid D)} = \frac{0.10}{0.90} \cdot \frac{256}{27} = \frac{256}{243} \approx 1.05$$

sabendo que $p(H_0\mid D) + p(H_1\mid D) = 1$, computamos que $p(H_0\mid D) = 0.487$ e concluimos que Tia Maude tem 48,7% de chance de estar falando verdade uma vez considerada a evidência.

Segundo caso: Tia Maude mente com 0,90 de chance.

$$\frac{p(H_0\mid D)}{p(H_1\mid D)} = \frac{0.90}{0.10} \cdot \frac{256}{27} = \frac{2304}{27} \approx 85.33$$

sabendo que $p(H_0\mid D) + p(H_1\mid D) = 1$, temos que a probabilidade a posteriori de Tia Maude estar falando verdade é $p(H_0\mid D) = 0.0116$.

3. **(A previsão do tempo em Metrópolis e Pequenópolis.)** Um canal de televisão faz a previsão diária do tempo dando 5 estimativas possíveis para o tempo naquele dia:

$$\mathcal{A}_X = \{\textit{ensolarado}, \textit{nublado}, \textit{chuvoso}, \textit{enevoado}, \textit{neve}\}\,.$$

O canal faz previsão para duas cidades, Metrópolis e Pequenópolis. Em Metrópolis, uma cidade fria e nebulosa, a frequência de dias com cada tipo de tempo é:

$$\mathcal{P}_X^{Metrópolis} = \left\{\frac{1}{10}, \frac{3}{10}, \frac{3}{10}, \frac{2}{10}, \frac{1}{10}\right\}.$$

Já em Pequenópolis, uma vila agradável e quentinha, a frequência de dias com cada tipo de tempo é:

$$\mathcal{P}_X^{Pequenópolis} = \left\{\frac{6}{10}, \frac{2}{10}, \frac{1}{10}, \frac{1}{10}, \frac{0}{10}\right\}.$$

1

(a) No dia 14 de setembro de 2017, o canal de televisão prevê tempo *ensolarado* tanto para Metrópolis quanto para Pequenópolis. Quanta informação, em bits, este resultado carrega para os habitantes de Metrópolis? E para os habitantes de Pequenópolis?

Em Metrópolis: $h(x = \text{ensolarado}) = \log \dfrac{1}{P_X^M(\text{ensolarado})} = \log \dfrac{1}{1/10} \approx \boxed{3.32 \text{ bits.}}$

Em Pequenópolis: $h(x = \text{ensolarado}) = \log \dfrac{1}{P_X^P(\text{ensolarado})} = \log \dfrac{1}{6/10} \approx \boxed{0.74 \text{ bits.}}$

2

(b) Seja $X^{Metrópolis} = (x, \mathcal{A}_X, \mathcal{P}^{Metrópolis})$ o ensemble representando o tempo em Metrópolis, e $X^{Pequenópolis} = (x, \mathcal{A}_X, \mathcal{P}^{Pequenópolis})$ o ensemble representando o tempo em Pequenópolis. Compute a entropia $H(X^{Metrópolis})$ e a entropia $H(X^{Pequenópolis})$.

$$H(X^M) = \frac{1}{10}\log 10 + \frac{3}{10}\log\frac{10}{3} + \frac{3}{10}\log\frac{10}{3} + \frac{2}{10}\log\frac{10}{2} + \frac{1}{10}\log 10$$

$$\approx \boxed{2.17 \text{ bits}}$$

$$H(X^P) = \frac{6}{10}\log\frac{10}{6} + \frac{2}{10}\log\frac{10}{2} + \frac{1}{10}\log 10 + \frac{1}{10}\log 10 + 0$$

$$\approx \boxed{1.57 \text{ bits.}}$$

2 (c) Usando as suas repostas para o item anterior (o cálculo de $H(X^{Metrópolis})$ e $H(X^{Pequenópolis})$), em qual cidade você diria que a previsão do tempo é mais útil? Por quê?

Como $H(X^M) \geq H(X^P)$, há mais incerteza sobre o tempo em Metrópolis do que em Pequenópolis. e, portanto, a previsão do tempo é mais útil em Metrópolis.

4. **(Compactação com perda.)** Seja um ensemble $X = (x, \mathcal{A}_X, \mathcal{P}_X)$ que representa a ocorrência de um botão apertado por um jogador de Playstation em uma partida online de *Mortal Kombat*, em que

$$\mathcal{A}_X = \{\square, \triangle, \bigcirc, \times, \text{R1}, \text{R2}, \text{L1}, \text{L1}, \text{options}, \text{share}, \text{Ps}\}$$

$$\mathcal{P}_X = \left\{\frac{15}{100}, \frac{20}{100}, \frac{10}{100}, \frac{20}{100}, \frac{10}{100}, \frac{5}{100}, \frac{10}{100}, \frac{5}{100}, \frac{2}{100}, \frac{2}{100}, \frac{1}{100}\right\}$$

Assuma que a rede é lenta, e que a sequência de botões apertadas por um jogador vai ser transmitida de forma compactada para o outro jogador online.

2 (a) Calcule o *raw-bit content* $H_0(X)$ da sequência de botões apertada pelo jogador nesta partida. Diga o que o valor encontrado significa.

$H_0(X) = \log |\mathcal{A}_X| = \log 11 \approx 3.46$ bits.

O valor $H_0(X)$ representa o número mínimo de bits necessário para dar um código de mesmo tamanho a cada símbolo de $\mathcal{A}_X$.

No caso, o código teria $\lceil H_0(X) \rceil = 4$ bits.

2 (b) Assuma que você quer compactar o ensemble X utilizando um código que tolera um erro de $\delta =$ na compactação e descompactação. Calcule o *essential bit content* $H_\delta(X)$ para $\delta = 0.15$, e diga o que o valor encontrado significa.

Primeiro computamos o conjunto S_δ para $\delta = 0.15$, que é $S_\delta = \{\square, \Delta, X, O, R1, L1\}$.

Então $H_\delta(X) = \log |S_\delta| = \log 6 \approx 2.58$ bits.

Este valor representa o número mínimo de bits necessários para dar um código de mesmo tamanho a cada símbolo de $\mathcal{A}_X$ se estamos dispostos a perder até $\delta = 0.15$ dos símbolos na codificação.

1 (c) Usando sua resposta no item (4b), crie um código binário de compactação para X com erro $\delta = 0.15$. Qual o tamanho esperado do código gerado?

$\square \mapsto 000$

$\Delta \mapsto 001$

$X \mapsto 010$

$O \mapsto 011$

$L1 \mapsto 100$

$R1 \mapsto 101$

O tamanho esperado é 3 bits por símbolo.

(d) (Não vale ponto, mas vale a admiração do professor.) Com o código com perdas que você gerou no item anterior, alguns botões que o jogador apertar não serão transmitidos na partida online para o outro jogador. Qual tipo de golpe especial do seu jogador favorito de *Mortal Kombat* (2011, X, ou XL) se tornaria impossível de ser executado nessa partida? Na sua opinião, se o custo de ter uma partida mais rápida for desabilitar esse tipo de golpe, essa compactação com perda vale a pena?

Não dá para fazer os golpes de X-Ray, que são os melhores! (Se não acredita em mim, cheque os golpes no Youtube!) :)