Aprendizagem Estatística
MAchine learning
learning from data

Seleçao Estatistica de modelos

Bibliografia

- Hastie, Tibshirani, Friedman. The elements of Statistical learning.
- Notas de Rafael.

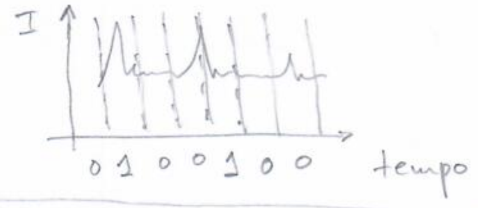
Objetivo: Queremos encontrar padroes em dados.

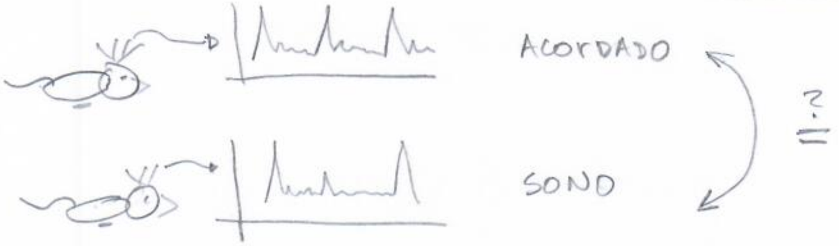
O que é um padrão ?

Exemplo 1: Um dos periodos do sono é REM (rapid eye movement)
Os registros REM parecem os registros feitos durante vigília
Sidarta Ribeiro (UFRN) tem a seguinte conjectura:

O cérebro "reverbera" as experiencias da vigilia durante o sono REM e assim aprende, constitui memória.

D. Brillinger, famoso estatístico dos anos 70 } Tive a ideia de discretizar uma sinai a sequencias de 0's e 1's





Exemplo 2 : (Seq 1)

Seq 1 : 0 1 1 0 1 0 0 0 1 0 1 0 0 1
Seq 2 : 0 1 0 0 1 0 0 0 1

tem um padrão comum ?

Ambas sequências têm proporção de números 1 ≈ 1/2
(talvez isso é um padrão)

$$\hat{p}_n \xrightarrow[n\to\infty]{} 1/2 \quad \text{Ley dos grandes números}$$

O que é um $\boxed{\text{padrão}}$???

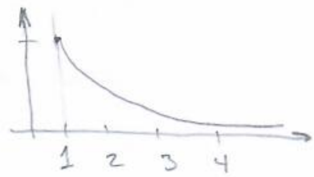
Exemplo 3: (Seq 2)

$$\underbrace{1\,0000\ldots0}\,111\ldots100\ldots01$$

C: comprimento é aleatorio e tem uma distribução geometrica de parametro p ($1/2 < p < 1$)

$$\mathbb{P}(C = k) = p^{k-1}(1-p)$$

$k = 1, 2, 3, \ldots$



$$\sum_{k=1}^{\infty} p^{k-1}(1-p) = \sum_{j=1}^{\infty} p^j (1-p)$$

$$= (1-p)\underbrace{\sum_{j=0}^{\infty} p^j}_{\text{progreção geométrica}}$$

$$= (1-p)\frac{1}{1-p}$$

$$= 1.$$

$$\underbrace{00000}_{C_1^0}\,\underbrace{1111\ldots1}_{C_1^1}\,\underbrace{000\ldots0}_{C_2^0}\,\underbrace{1\ldots1}_{C_2^1}$$

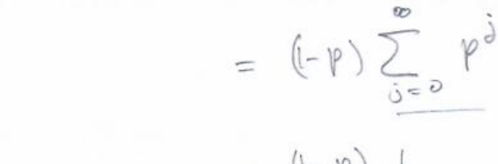
$$\mathbb{P}(C_i^0 = k) = p^{k-1}(1-p)$$

$$\mathbb{P}(C_i^1 = k) = p^{k-1}(1-p)$$

Generador Seq 1

1. Escolo uma bola:
2. Si a bola é bramca (O) coloco 0.
   Si a bola é amarela (□) coloco 1.

Generador Seq 2



Urna 0 Urna 1

1. Escolo uma das 2 urnas ao acaso com prob. 1/2, 1/2
2. Escolo ao acaso uma bola da urna sorteada. Se a bola for da cor majoritaria recomeço na urna

e faço novo sorteio

3. Se a bola for da cor minoritaria, mudo de urna
4. Depois de cada sorteio reponho a bola na urna.

Exercício (para casa)

Escrever pseudo-códigos para implementar algoritmos gerando as 2 experiencias de urna; mais a seguinte experiencia

Jogo do Goleiro:

211 211 211 211

$\underbrace{\quad}_{1^{ro} \text{ ou } 1^{ro} \text{ ou } 1^{ro}}$

1/3 1/3 1/3

$\to 1 \to$ 1 com prob 0.8

$\to 1 \to$ 0 com prob 0.2

em cada símbolo 1 independente um dos outros

Geração

1. Escolho começar com 2
   ou com 1^{ro} simbolo 1 com probabilidade 1/3, 1/3, 1/3
   ou com 2^{do} simbolo 1
2. A partir do simbolo escolhido concateno
3. Atualizo cada simbolo 1 de maneira iid da seguinte maneira:
   1 permanece 1 com prob 0.8
   1 é alterado para 0 com prob. 0.2

M. Gromov (Geómetra)

Padrão: (Definição provisoria)

Conjunto coherente de regularidade estatística.

Resumo: O qué fizemos na 1^{ra} parte.

- Discutimos como "classificar" sequências de 0's e 1's.
- Tentamos fazer isso utilizando proporçoes relativas
- No final atribuimos um algoritmo de geração para cada sequencia.

Sequencia $x_1 x_2 \dots x_n = x$, $x_i \in \{0,1\}$

$\downarrow$

modelo - [algoritmo de geração

Problema de clasificaçao sem supervisão

Clasificação com supervisão:

$\mathcal{S}$: conjunto

$\xi_1, \xi_2, \dots, \xi_n$ são elementos de $\mathcal{S}$

f

$y_1, y_2, \dots, y_n$, $y_n = f(\xi_n)$, $y_n \in \{0,1\}$

(ξ_n, y_n) são iid

Hipotese simplificadora: y_n = funçao determinista de ξ_n

Dado amostra $(\xi_1, y_1), \dots (\xi_n, y_n)$ iid.
Sei que y_n = função (ξ_n) e sei tambem que ea função desconhecida pertence ao conjunto $\mathcal{H}$ de funções

$\mathcal{H}$ = conjunto de "hipóteses" ou de "modelos" que tenho à minha disposição

Problema: Escolher $\hat{f}_n \in \mathcal{H}$ que "melhor se ajuste" aos dados

Erro do ajuste (Risco):

Tomo $h \in \mathcal{H}$,

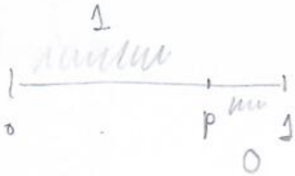
$$R(h) = \text{probabilidade de } h(\xi) \neq y$$
$$= \mathbb{P}(h(\xi) \neq y)$$

$\hookrightarrow$ não conheço $\mathbb{P}$.

Obs:

$f = \arg\min \{R(h) : h \in \mathcal{H}\}$, $R(f) = 0$ por hipotese

Exemplo:



$\mathcal{S} = [0,1]$

$\xi_1, \xi_2, \dots, \xi_n$ são sorteios independentes feitos com uma distribuição uniforme em $[0,1]$.

$y_n = \mathbb{1}_{\{\xi_n \leq p\}}$ onde $0 < p < 1$ fixado.

$$\mathcal{H} = \left\{ h : [0,1] \to \{0,1\} : h \text{ é da forma } h(u) = \mathbb{1}_{\{u \leq q\}} \text{ para algum } q \in [0,1] \right\}$$

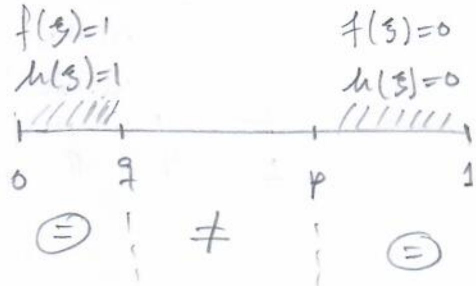
$\mathbb{P}$ = distribuição uniforme em $[0,1]$

$R(h) = \mathbb{P}\left(h(\xi) \neq f(\xi) \right)$ onde $h(\xi) = \mathbb{1}_{\{\xi \leq q\}}$

$$\mathbb{P}\left([q,p] \right) = p - q$$

Vamos supor que $q < p$

$$f(\xi) = \begin{cases} 1 & \text{se } \xi \leq p \\ 0 & \text{se } \xi > p \end{cases}$$



Em geral $R\left(\mathbb{1}_{\{\xi \leq q\}} \right) = |p - q|$, e $\min h(h) = 0$ ocorre quando $p = q$.

Escolho $h \in \mathcal{H}$, isto é h da forma $\mathbb{1}_{[0,q[}$ para algum $q \in [0,1]$

$$\hat{R}_n(h) = \frac{1}{n} \sum_{m=1}^{n} \mathbb{1}_{\{h(\xi_m) \neq y_m\}} \quad \text{onde } (\xi_1, y_1), \dots, (\xi_m, y_m)$$

é a amostra de que dispomos

$$\hat{R}_n(h) \xrightarrow[n \to \infty]{} R(h) \quad \Big| \text{ Ley dos grandes números}$$

Fixo δ,

$$\mathbb{P}\left(\left| \hat{R}_n(h) - R(h) \right| > \delta \right) \leq \underbrace{\varepsilon(n)}_{} \downarrow 0$$

gostaria de ter a menor majoração possível.

$$\hat{R}_n(h) - R(h)$$

$$\underbrace{\frac{1}{n} \sum_{m=1}^{n} \underbrace{\mathbb{1}_{\{h(\xi_m) \neq y_m\}}}_{Z_m}}_{\mathbb{E}\left(\mathbb{1}_{\{h(\xi) \neq y\}}\right)} \qquad \mathbb{P}\left(h(\xi) \neq y\right) \quad \text{onde } y = f(\xi)$$

$Z_m \longrightarrow$ var. aleatorias iid

Obs.:

$A \subset [0,1]$

$\mathbb{P}(A) = \mathbb{E}(\mathbb{1}_A)$

$$\mathbb{1}_A = \begin{cases} 1 & \text{se } u \in A \\ 0 & \text{se } u \notin A \end{cases}$$

$$\mathbb{E}(\mathbb{1}_A) = 1 \cdot \mathbb{P}(u : u \in A) + 0 \cdot \mathbb{P}(u : u \notin A) = \mathbb{P}(A)$$

$$\hat{R}_n(h) = \frac{1}{n} \sum_{m=1}^{n} Z_m, \quad R(h) = \mathbb{E}(Z)$$

$$\hat{R}_n(h) - R(h) = \underbrace{\frac{1}{n} \sum_{m=1}^{n} Z_m}_{\substack{\text{media} \\ \text{empirica} \\ \text{(conheço)}}} - \underbrace{\mathbb{E}(Z)}_{\substack{\text{media teórica} \\ \text{(não conheço)}}}$$

Lei dos grandes números diz que sob certas hipoteses a media empírica converge para a media teórica.

$$\mathbb{P}\left(|\hat{R}_n(h) - R(h)| > \delta\right) = \mathbb{P}\left(\left|\frac{1}{n}\sum_{m=1}^{n} z_m - \mathbb{E}(z)\right| > \delta\right)$$

$$= \mathbb{P}\left(\left|\sum_{m=1}^{n} z_m - n\,\mathbb{E}(z)\right| > n\delta\right)$$

$$\leq \underbrace{\frac{1}{n}\,\frac{var(z)}{\delta^2}}_{=\,\varepsilon(n)}$$

Desigualdade de Chebyshev

esto pode ser majorado por 1/4

Na verdade podemos obter:

$$\varepsilon(n) = e^{-2n\delta^2}$$

Desigualdade de Hoeffding (1960).