Sumário do Capítulo 2

Definição: A *entropia* de uma v.a. discreta é definida por

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

Propriedades de H

1. $H(X) \geq 0$.
2. $H_b(X) = (\log_b a) H_a(X)$
3. (Condicionamento reduz a entropia) Dadas duas v.a.s X e Y, tem-se

   $$H(X|Y) \leq H(X)$$

   com igualdade sse X e Y forem independentes.
4. $H(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i)$, com igualdade sse as v.a.s X_i forem independentes.
5. $H(X) \leq \log |\mathcal{X}|$, com igualdade sse X é uniformemente distribuído em $\mathcal{X}$.
6. $H(p)$ é concava em p.

Definição: A *entropia relativa* $D(p\|q)$ entre as distribuições p e q é definida como

$$D(p\|q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}.$$

Definição: A *informação mútua* $I(X;Y)$ entre as v.a.s X e Y é definida como

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}.$$

Expressões alternativas:

$$\begin{aligned}
H(X) &= -E_p \log p(X) \\
H(X,Y) &= -E_p \log p(X,Y) \\
H(X|Y) &= -E_p \log p(X|Y) \\
I(X;Y) &= E_p \log \frac{p(X,Y)}{p(X)p(Y)} \\
D(p\|q) &= E_p \log \frac{p(X)}{q(X)}
\end{aligned}$$

Propriedades de D e I

1. $I(X;Y) = H(X) - H(X|Y) = I(Y;X) = H(X) + H(Y) - H(X,Y)$.
2. $D(p\|q) \geq 0$ com igualdade sse $p(x) = q(x)$, para todo $x \in \mathcal{X}$.
3. $I(X;Y) = D(p(x,y)\|p(x)p(y))$, com igualdade sse $p(x,y) = p(x)p(y)$, i.e., X e Y são independentes.

Regras da cadeia:

Entropia:	$H(X_1, ..., X_n) = \sum_{i=1}^{n} H(X_i \| X_i - 1, \ldots, X_i)$
Informação Mútua:	$I(X_1, ..., X_n; Y) = \sum_{i=1}^{n} I(X_i; Y \| X_i - 1, \ldots, X_i)$
Entropia relativa:	$D(p(x,y)\|q(x,y)) = D(p(x)\|q(x)) + D(p(y\|x)\|q(y\|x))$.

Desigualdade de Jensen: Qualquer função convexa f verifica $Ef(X) \geq f(EX)$

Desigualdade de Gibbs: Dadas duas distribuições de probabilidade p e q, tem-se

$$\sum_i p_i \log \frac{q_i}{p_i} \leq 0,$$

com igualdade sse $p_i = q_i$ para todo i.

Cadeias de Markov: Dada a cadeia de Markov $X \to Y \to Z$, tem-se

$$\begin{aligned} I(X;Y) &\geq I(X;Z) \\ I(X;Y) &\geq I(X;Y|Z). \end{aligned}$$

Desigualdade de processamento de dados: Dadas as v.a.s X e Y tem-se sempre

$$I(X;Y) \geq I(X;g(Y)).$$