Teoria da Informação

Charles Casimiro Cavalcante

charles@gtel.ufc.br

Grupo de Pesquisa em Telecomunicações Sem Fio – GTEL
Programa de Pós-Graduação em Engenharia de Teleinformática
Universidade Federal do Ceará – UFC
http://www.gtel.ufc.br/∼charles

"A principal função de um sistema de comunicação é reproduzir, exatamente ou de forma aproximada, uma informação proveniente de outro ponto diferente."

Claude Shannon, 1948

Conteúdo do curso

1. Revisão de probabilidade
2. Informação e Entropia
3. Codificação de fontes
4. Codificação e capacidade de canal
5. Complexidade de Kolmogorov
6. Funções de otimização
7. *Independent Component Analysis*

Parte VI

Funções de Otimização

Considerações

- Funções de otimização são utilizadas para atingir determinado objetivo em um processo físico, descrito através de uma formulação matemática
- Funções de otimização são, principalmente, ligadas à questões relacionadas aos problemas de *controle* e **filtragem**
- Funções de otimização são também chamadas de funções objetivo ou funções custo

Considerações - cont.

- Em sistemas de comunicação, conforme dito por Shannon, a comunicação tem por meta reproduzir da melhor maneira possível no receptor uma informação enviada pelo transmissor
- Como medir esta "confiabilidade de transmissão"?
- Shannon, ainda nos anos 30-40 propôs a seguinte medida

$$\min \ \mathcal{E} = \int_{-\infty}^{\infty} \left(a(t) - \widehat{a}(t)\right)^2 \ dt \tag{121}$$

Considerações - cont.

- A idéia é que minimizando-se o erro quadrático, prova-se que a estimativa do sinal tende, assintoticamente, para o valor original
- Resultado aproveitado por Wiener posteriormente (1948)
- Base da maioria de técnicas de filtragem e controle

Problema

Como calcular a estimativa do sinal?

Teoria da estimação

- A maneira de estimar o sinal depende do processo empregado
- Otimização/filtragem **linear** calculam a estimativa através de uma combinação linear de parâmetros

$$\widehat{a}(t) = \sum_{k=-\infty}^{\infty} w_k \cdot x(k) \tag{122}$$

- Otimização/filtragem **não-linear** utilizam funções mais complexas para estimativa do sinal

$$\widehat{a}(t) = \sum_{k=-\infty}^{\infty} f(w_k) \cdot g[x(k)] \tag{123}$$

- Método dos momentos
- Métodos dos mínimos quadrados (linear, generalizado e não-linear)
- Método da máxima verossimilhança
- Método bayesiano

Ainda problema...

Estimativa do sinal apresenta uma forte dependência do tipo de processamento utilizado. Como proceder para se obter uma estimativa que independe do tipo de abordagem (linear ou não-linear)?

Teoria da informação × teoria da estimação

- Resposta: abordagem pela transferência de informação!!
- Ferramentas: medida de informação
- Otimização por critérios baseados em teoria da informação (Information-theoretic criteria)
- Aplicação em vários problemas de engenharia

Medidas de otimização

- **Maximização da informação mútua**

$$J = \max\ \mathcal{I}(X, Y)$$

  Aplicação direta em sistemas discretos

- **Maximização da entropia**

$$J = \max\ \mathcal{H}(Y)$$

  Busca por uma densidade de maior energia

Medidas de otimização - cont.

- **Maximização da negentropia**

$$J = \max \ (\mathcal{H}(Y_G) - \mathcal{H}(Y))$$

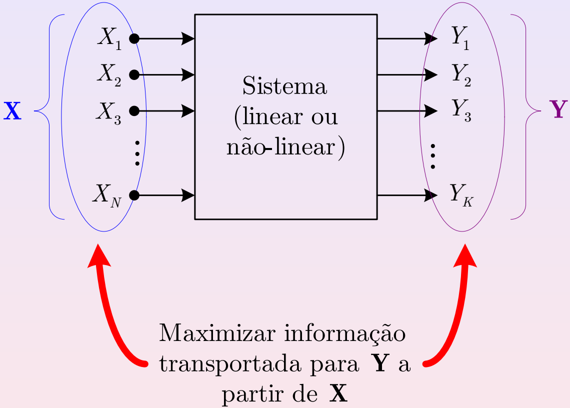
  Para uma variância dada, torna os dados o mais não-gaussiano posssível

- **Maximização/minimização da divergência de Kullback-Leibler**

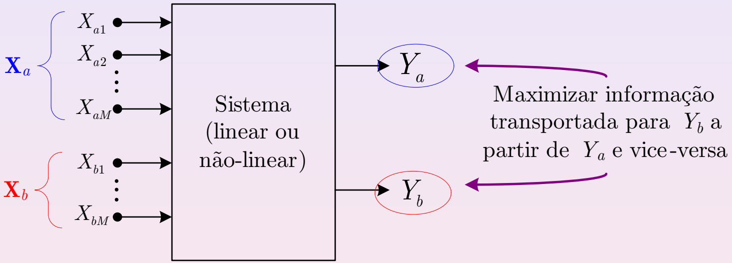
$$J = \max \ D(p||g)$$

  Maximizar/minimizar a similaridade entre duas funções definidas positivas

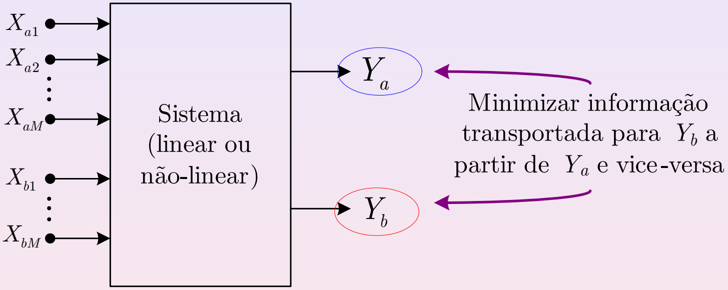
- Em particular, a medida de informação mútua é interessante na aplicação da determinação de funções obetivo pois esta pode ser implementada de forma não-supervisionada
- Quatro tipos básicos de cenário, cuja escolha depende da aplicação de interesse
  1. Maximizar a quantidade de informação transportada para a saída do sistema pela entrada do mesmo
  2. Maximizar a quantidade de informação transportada entre partes da saída do sistema
  3. Minimizar a quantidade de informação transportada entre partes da saída do sistema
  4. Minimizar a dependência estatística entre as saídas de um sistema



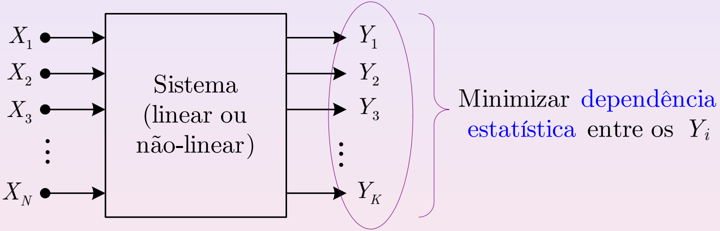
Maximização da informação (capacidade de canal!)



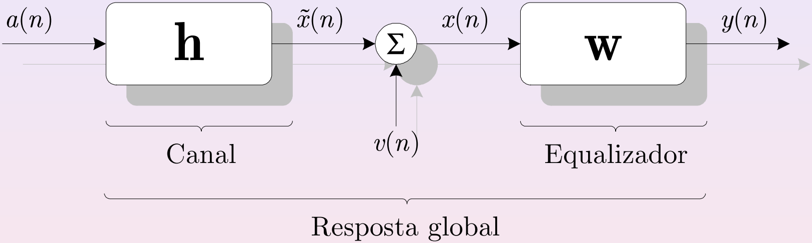
Partes diferentes de uma mesma imagem



Partes de imagens diferentes



Recuperação de informação da entrada



$a(n)$
$\mathbf{h}$
$\tilde{x}(n)$
Σ
$x(n)$
$\mathbf{w}$
$y(n)$
$v(n)$
Canal
Equalizador
Resposta global

Aspectos gerais

- Proposição inicial para sistemas SISO (equalização mono-usuário)
- Caracterização da função de densidade de probabilidade (fdp) da saída do *equalizador ideal*
- Estimação paramétrica
- Critério: estimação/entropia
  - Divergência de Kullback-Leibler (KLD)
  - Função contraste

Sinal na saída do equalizador ideal

$$
\begin{aligned}
y(n) &= \left(\mathcal{H}^H \mathbf{a}(n) + \mathbf{v}(n)\right)^H \mathbf{w}_{\text{ideal}} \\
&= \mathbf{a}^H(n)\mathcal{H}\mathbf{w}_{\text{ideal}} + \mathbf{v}^H(n)\mathbf{w}_{\text{ideal}} \\
&= \mathbf{a}^T(n)\underbrace{\mathcal{H}\mathbf{w}_{\text{ideal}}}_{\mathbf{g}_{\text{ideal}}} + \mathbf{v}^H(n)\mathbf{w}_{\text{ideal}} \\
&= \mathbf{a}^H(n)\mathbf{g}_{\text{ideal}} + \vartheta(n) \\
&= a(n-\ell) + \vartheta(n),
\end{aligned}
$$

pdf de $y(n)$

$$
p_Y(y) = \frac{1}{\sqrt{2\pi\sigma_\vartheta^2}} \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n) - \mathfrak{a}_i|^2}{2\sigma_\vartheta^2}\right) \Pr(\mathfrak{a}_i),
$$

Modelo paramétrico

$$\Phi(y) = \underbrace{\frac{1}{\sqrt{2\pi\sigma_r^2}}}_{A} \cdot \frac{1}{\mathfrak{C}} \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n) - \mathfrak{a}_i|^2}{2\sigma_r^2}\right), \quad (124)$$

Como medir a similaridade?

Divergência de Kullback-Leibler (KLD)

$$D\left(p_Y(y) || \Phi(y)\right) = \int_{-\infty}^{\infty} p_Y(y) \cdot \ln\left[\frac{p_Y(y)}{\Phi(y)}\right] dy$$

Fitting pdf Criterion

Termo da KLD dependente de $\Phi(y)$

$$\begin{aligned}
J_{\mathsf{FP}}(\mathbf{w}) &= \int_{-\infty}^{\infty} p_Y(y) \cdot \ln\left(\frac{1}{\Phi(y)}\right) dy \\
&= -\mathbb{E}\left\{\ln\left[\Phi(y)\right]\right\} \\
&= -\mathbb{E}\left\{\ln\left[A \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y-\mathfrak{a}_i|^2}{2\sigma_r^2}\right)\right]\right\}
\end{aligned}$$

Fitting pdf Algorithm

$$\nabla J_{\mathsf{FPC}}(\mathbf{w}) = \frac{\sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n)-\mathfrak{a}_i|^2}{2\sigma_r^2}\right)[y(n)-\mathfrak{a}_i^*]}{\sigma_r^2 \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n)-\mathfrak{a}_i|^2}{2\sigma_r^2}\right)} \mathbf{x}(n),$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla J_{\mathsf{FPC}}(\mathbf{w}).$$

Equivalências

- Solução do FPA quando se dispõe dos dados coincide com a solução MMSE
- Pode ser visto como um caso generalizado dos critérios CM e de Sato
- Equivalência do comportamento da função custo com algoritmo de decisão dirigida
- Relação importante com critério de minimização da probabilidade de erro

FP × MAP

$$J_{\text{MAP}} = J_{\text{FPC}} - J_{\text{MMSE}}$$