

## Pré e Pós-processamento dos dados

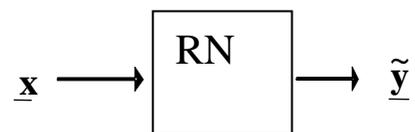
### Preparação dos dados de entrada para treinamento não supervisionado

(parecido – mas não igual – ao das redes feedforward)

- 1 - Escolha das variáveis de entrada
- 2 – Compactação / Parametrização das variáveis
- 3 – Escalamento das variáveis
- 4 - Pares entrada – saída
- 5 – Relevância das entradas

#### 1 - Escolha das Variáveis de Entrada

“Como entrada escolha as variáveis relevantes, todas as variáveis relevantes e somente as variáveis relevantes **para a classificação desejada**.”



Usualmente a classificação depende fortemente das entradas utilizadas

#### Como saber se uma entrada é relevante ?

Fenomenologia >>>> candidatas à relevantes

Análise estatística (pré-processamento)

Relevância (pós-processamento)

**Entradas não relevantes dificultam o treinamento (porque se comportam como ruído) e eventualmente podem gerar novas classes “artificiais”, sem interesse.**

## 1.1 - Treinamento Supervisionado

### Independência / Dependência Estatística entre Variáveis

#### Coefficiente de Correlação de Pearson

Pares  $(x_i, y_i)$   $i = 1, \dots, P$

$$r(x, y) = \frac{\frac{1}{P-1} \sum_{i=1}^P (x_i - \mu_x)(y_i - \mu_y)}{\sigma_x \sigma_y} \quad \text{onde}$$

$$\mu_x = \frac{1}{P} \sum_{i=1}^P x_i \quad e \quad \sigma_x = \sqrt{\frac{1}{P-1} \sum_{i=1}^P (x_i - \mu_x)^2}$$

**embora não seja o mais recomendado, também oferece informação relevante usado com variáveis discretas.**

**95% confiança na correlação**

$$|r| \geq 2\sigma(r) = \frac{2}{\sqrt{P}}$$

#### Matriz de correlações entradas – saídas

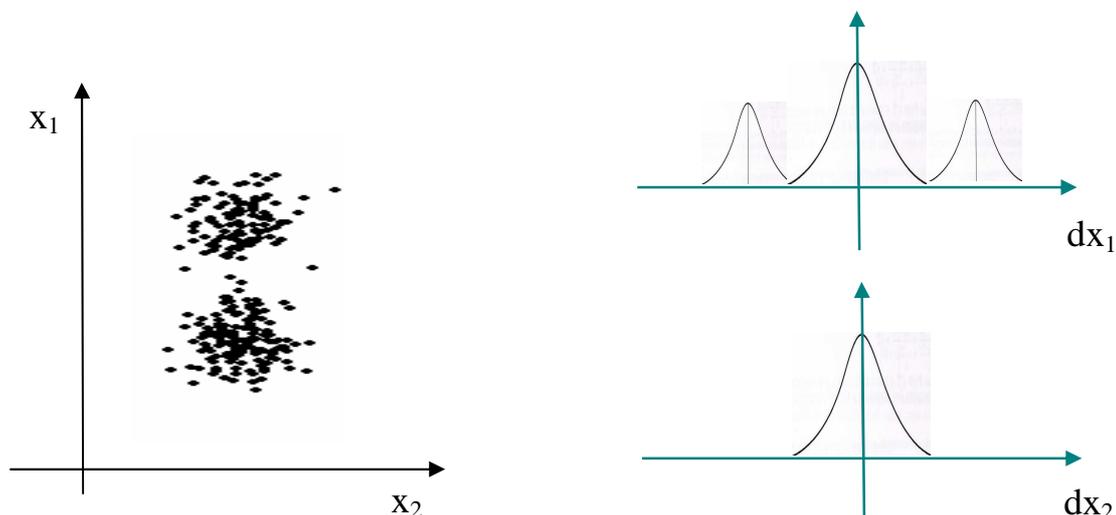
	$y_1$	...	$y_m$
$x_1$	$r_{1y1}$		$r_{1ym}$
$x_2$	$r_{2y1}$		$r_{2ym}$
$x_3$	$r_{3y1}$		$r_{3ym}$
...			
$x_n$	$r_{ny1}$		$r_{nym}$

**usada para decidir se a variável será utilizada ou não**

## 1.2 - Treinamento não supervisionado

### Analisar a distribuição das variáveis em cada dimensão

Variáveis com modas bem diferenciadas no histograma de diferença de valores entre os elementos tem melhor possibilidade de gerar bons agrupamentos.



## 2 – Compactação / Parametrização das variáveis

**Informação redundante:**

**Ex: Voz, Imagens, Sonar, etc.**

**pode reduzir o ruído mas**

**dificulta e torna lento o treinamento e a operação**

**Solução: Compactar e/ou parametrizar entradas muito redundantes.**

**Mas a parametrização pode alterar a classificação.**

## 2.1 Processos de parametrização / compactação:

### – Baseados na Fenomenologia

**Ex:** Formantes e coeficientes cepstrais para voz,  
Parâmetros de taxonomia biológica, etc.

### - Transformadas Matemáticas:

**Ex:** Fourier, Wavelets, QV, etc.  
PCA, PCA generalizadas, ICA, etc.

### – Processos de Invariância, Insensibilidade

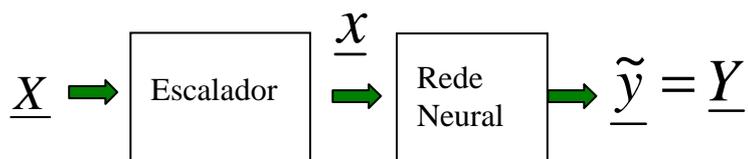
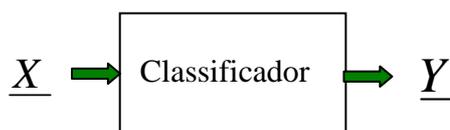
**Ex:** Invariância à translação, escala e rotação de imagens  
Insensibilidade ao locutor na análise de conteúdo de voz, etc.

## 3 – Escalamento das Variáveis

fundamental para o bom condicionamento do processo numérico

$X_i =$  variável original  $\gggg$   $x_i =$  variável escalada

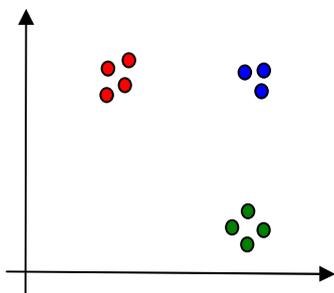
Treinamento e Operação da Rede com variáveis normalizadas



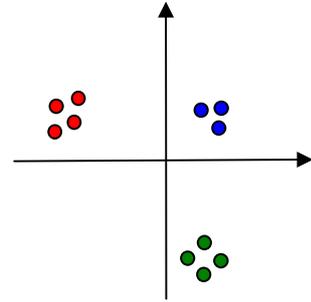
### 3.1 – Invariâncias e dependências

#### Agrupamento “natural”:

Insensível à translação



$$\underline{x} + \underline{b}$$



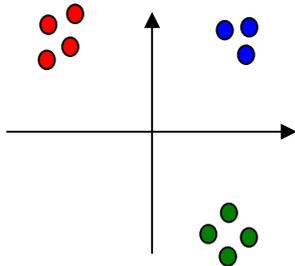
Usar variáveis com média zero

$$\underline{x} \implies \underline{x} - \underline{\mu}_x$$

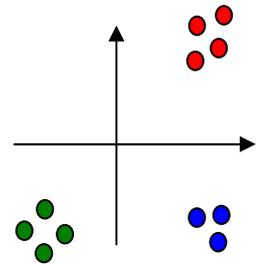
$$x_i \implies x_i - \mu_{xi}$$

#### Agrupamento “natural”:

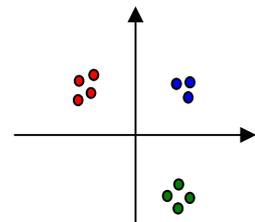
Insensível à

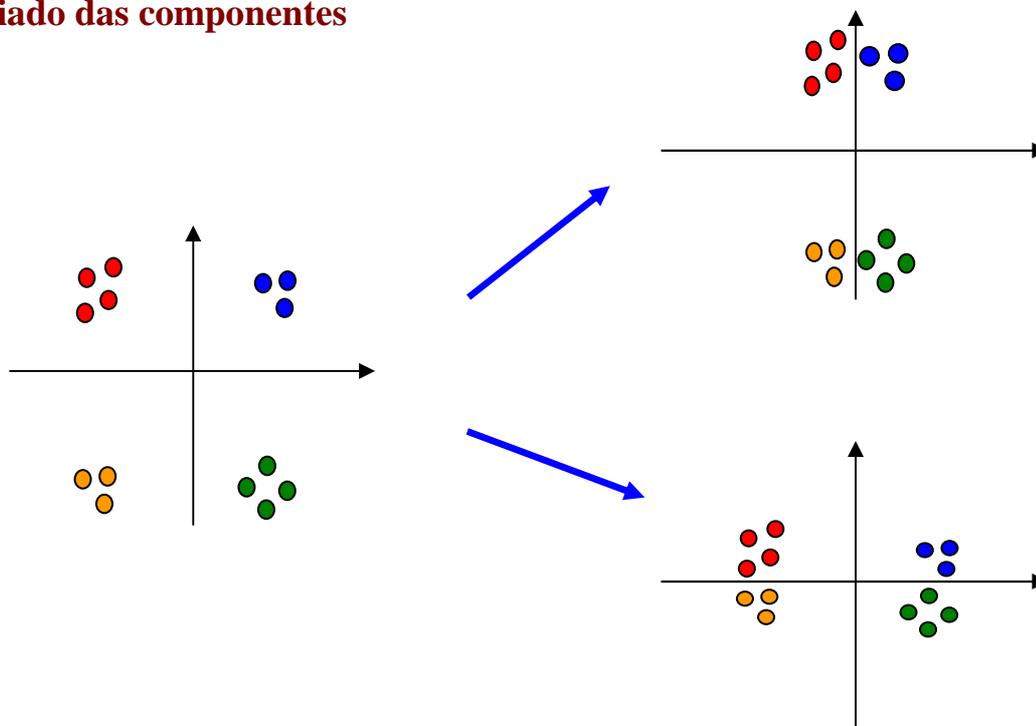


Rotação  
 $\underline{R} \underline{x}$



Escala total  
 $k \underline{x}$



**Sensível ao escalamento****diferenciado das componentes****3.2 - Variáveis quantitativas (entradas):**

**contínuas** (e.g. temperatura, comprimento)

**discretas** (e.g. número de filhos) – representável por variável contínua

usar média nula e

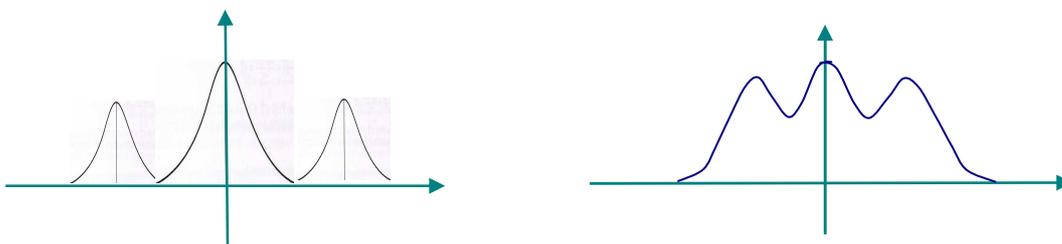
dispersão **adequada** por variável:

$$x_i = k_i (X_i - \mu_{X_i})$$

$x_i$  tem média nula e  $k_i$  controla sua dispersão.

### 3.2.1 - Controle da dispersão $k_i$

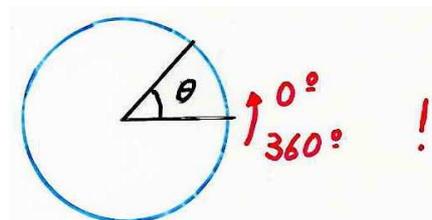
Escolher  $k_i$  tal que o desvio padrão intra classe em toda direção unitário.  
Ver cap ART Classes



$$k_i = \frac{1}{\sigma_i} \approx \frac{\sqrt{2}}{\sigma_d}$$

### 3.2.3 Variáveis cíclicas

Como para o MLP  
Não introduzir descontinuidades abruptas  
em variáveis originalmente contínuas



$$\theta \Rightarrow \text{sen } \theta, \text{ cos } \theta$$

$$\mathbf{X: 0 - 24 h.} \quad \ggggg \quad \mathbf{x: ( \text{sen } 2\pi X/24 ; \text{cos } 2\pi X/24 )}$$

$$\mathbf{X: 0 - 12 meses} \quad \ggggg \quad \mathbf{x: ( \text{sen } 2\pi X/12 ; \text{cos } 2\pi X/12 )}$$

### 3.2.4 Variáveis cobrindo faixas muito extensas (várias décadas)

Como para o MLP

Escalamento não linear: as variáveis podem ser comprimidas  
em uma escala logarítmica antes da normalização.

Note que utilizar log pode modificar os agrupamentos.

### 3.3 Variáveis categóricas (entradas e saídas)

**binárias** ( e.g.  $X_i \in \{\text{frio, quente}\}$  ou  $X_i \in \{\text{feio, bonito}\}$ )

$$x_i \in \{0,1\}$$

**nominais** ( e.g.  $X_i \in \{\text{solteiro, casado, separado, viúvo}\}$ )

$x_i$  em notação binária  
maximamente esparsa

$$\text{e.g. } \underline{x}_i \in \left\{ \begin{array}{c} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \end{array} \right\}$$

Ao contrário das redes MLP, aqui não há vantagens em se trabalhar com  $\{-1,+1\}$

Variáveis categóricas podem influir fortemente na classificação.

#### 4 - Pares entrada – saída:

##### População por classe

Normalmente pouco crítico, mas para alguns processos classes com pouca população podem não ser aprendidas corretamente

##### Intrusos (outlayers, outsiders)

Muito menos críticos que na backpropagation. Para a maioria dos processos simplesmente aparecem como uma classe com um único elemento.

##### Pares com componentes faltando

Evitar o uso. Caso imprescindível, fazer o treinamento / operação substituindo o componente pelo valor médio (zero), e eventualmente afrouxando o raio de similaridade.

## 5 – Relevância das entradas (pós-processamento)

em um classificador uma entrada é muito relevante se sua ausência provoca um percentual elevado de modificações nas saídas (na classificação) para o conjunto das  $P$  entradas disponíveis.

**relevância de uma entrada  $x_i$  em relação à saída  $y_j$  (à classe  $C_j$ )**

$$\Delta x_i \Rightarrow \Delta y_j \begin{cases} \Delta y_j \gg & \text{muito relevante} \\ \Delta y_j \ll & \text{pouco relevante} \end{cases}$$

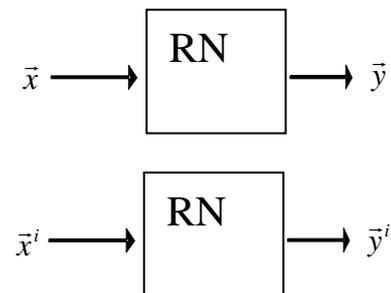
### 5.1 Medida da relevância

considere

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_i \\ \dots \\ x_n \end{bmatrix} \quad e \quad \bar{x}^i = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ \mu(x_i) \\ \dots \\ x_n \end{bmatrix} \quad \text{onde} \quad \mu(x_i) = \begin{cases} 0 & \text{se } x_i \text{ é variável contínua} \\ 0,5 & \text{se } x_i \in \{0,1\} \text{ é variável lógica} \end{cases}$$

aplicando  $\bar{x}$  e  $\bar{x}^i$  na rede obteremos respectivamente  $\bar{y}$  e  $\bar{y}^i$

$$\text{onde} \quad \bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_j \\ \dots \\ y_m \end{bmatrix} \quad e \quad \bar{y}^i = \begin{bmatrix} y_1^i \\ y_2^i \\ \dots \\ y_j^i \\ \dots \\ y_m^i \end{bmatrix}$$



realizando esta operação para  $P$  entradas,

$$r(x_i, y_j) = \frac{1}{P} \sum_{\forall P \text{ entradas}} (y_j - y_j^i)^2$$

$r(x_i, y_j)$  é a relevância de  $x_i$  em relação à  $y_j$  e mede o percentual de vezes em que a classe  $C_j$  é afetada pela eliminação da entrada  $x_i$

$$r(x_i) = \sum_{j=1}^m r(x_i, y_j) = \frac{1}{2P} \sum_{\forall P \text{ entradas}} \sum_{j=1}^m (y_j - y_j^i)^2$$

$r(x_i)$  é a relevância total de  $x_i$  em relação à classificação, e mede o percentual de vezes em que a classificação é afetada pela eliminação da entrada  $x_i$ .

Obs: o fator 2 em  $r(x_i)$  é porque devido ao winner-takes-all mudam 0 ou 2 saídas.