

Temas PSI3471 da quarta semana c/ Prof. Emilio

#7 (23/março – 2ªf) Foco da semana: Reconhecedores – Circuitos lógicos genéricos e suas implementações através de computação neural; relação com discriminadores. Construindo reconhecedores de padrões com fronteiras de separação genéricas: separadores lineares; bancos de separadores lineares como aproximadores efetivos de fronteiras genéricas.

#8 (25/março – 4ªf) ... Técnicas de medida de qualidade em reconhecedores de padrões: matriz de confusão, conceitos de falsos positivo e negativo, especificidade e sensibilidade. Curvas ROC.

© Prof. Emilio Del Moral Hernandez

1

Emílio Del Moral Hernandez

Slides comentados
Prof. Emilio Del Moral Hernandez

Site do Grupo ICONE — <http://www.lsi.usp.br/ICONE/>
... e facebook: www.facebook.com/ICONE.EPUSP/

Grupo de Inteligência Computacional, Modelagem e Neurocomputação - ICONE
Laboratório de Sistemas Integrados - LSI
Escola Politécnica da USP - EPUSP
Universidade de São Paulo

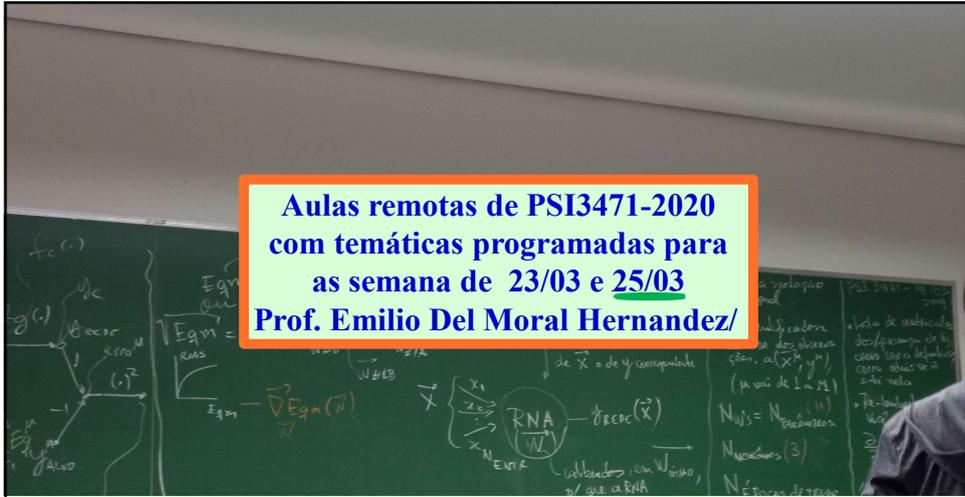
XXI WACONE
Matriz 21 de Setembro de 2017
Resumo dos trabalhos apresentados
Local: Auditorio de Edifício

XXII WACONE
Matriz 21 de Setembro de 2017
Resumo dos trabalhos apresentados
Local: Auditorio de Edifício

XXIII WACONE
Matriz 21 de Setembro de 2017
Resumo dos trabalhos apresentados
Local: Auditorio de Edifício

18:22 Role para ver detalhes
© Prof. Emilio Del Moral Hernandez

127



**Aulas remotas de PSI3471-2020
com temáticas programadas para
as semana de 23/03 e 25/03
Prof. Emilio Del Moral Hernandez/**

Temas da quarta semana c/ Prof Emilio

#7 (23/março – 2ªf) Foco da semana: Reconhecedores – Circuitos lógicos genéricos e suas implementações através de computação neural; relação com discriminadores. Construindo reconhecedores de padrões com fronteiras de separação genéricas: separadores lineares; bancos de separadores lineares como aproximadores efetivos de fronteiras genéricas.

#8 (25/março – 4ªf) ... Técnicas de medida de qualidade em reconhecedores de padrões: matriz de confusão, conceitos de falsos positivo e negativo, especificidade e sensibilidade. Curvas ROC.

... nestes slides: segunda destas 2 aulas (#8 -25/03)

Temas PSI3471 da quarta semana c/ Prof Emilio

#7 (23/março – 2ªf) Foco da semana: Reconhecedores – Circuitos lógicos genéricos e suas implementações através de computação neural; relação com discriminadores. Construindo reconhecedores de padrões com fronteiras de separação genéricas: separadores lineares; bancos de separadores lineares como aproximadores efetivos de fronteiras genéricas.

#8 (25/março – 4ªf) ... Técnicas de medida de qualidade em reconhecedores de padrões: matriz de confusão, conceitos de falsos positivo e negativo, especificidade e sensibilidade. Curvas ROC.

*Medidas de desempenho
em reconhecedores de padrões /
detectores e
classificadores*

© - Prof. Emilio Del Moral Hernandez

133

**Algumas formas de medir a qualidade de
detectores / reconhecedores de padrões?**

- Taxa de acertos (%)
- Taxa de erros (% complementar)
- Taxa de acertos em separado para as duas classes
(A x B) ... Taxas de acerto / erro para A .. e para B ...

Sensibilidade, Especificidade, etc etc

... Sensibilidade = $VP / \text{Todos casos realmente Positivos}$

... Especificidade = $VN / \text{Todos casos realmente Negativos}$

134

© - Prof. Emilio Del Moral Hernandez

134

Continuando com + conceitos: avaliando o desempenho de reconhecedores de padrões

- O Erro quadrático médio E_{qm} (ou o RMSE, sua raiz quadrada), que é fornecido automaticamente pelas bibliotecas neurais ... poderia até ser usado, mas há indicadores muito mais interessantes, que são específicos para o contexto de reconhecimento e classificação de padrões, e são mais legítimos e informativos que o E_{qm} / RMSE
- Taxa de acertos em classificação / reconhecimento
- 4 Combinações possíveis (ou $(K_{classes})^2$ no caso geral multiclases) envolvendo as classificações do classificador “padrão ouro” e as classificações do modelo neural
(diagramas de Venn / conjuntos nos ajudam a entender)
- Matriz de confusão: tamanho é $K_{classes} \times K_{classes}$

137

© - Prof. Emilio Del Moral Hernandez

137

(I) A imperfeição na separabilidade de classes (taxa de acertos menos que 100%) pode ser gerada por limitações da fronteira separadora da classe de modelagem empregada:

fronteira linear

x

não linear (“curva parametrizada”)

x

fronteira de formato qualquer

138

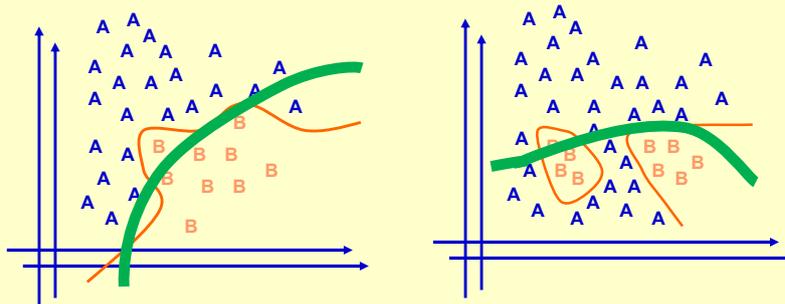
© - Prof. Emilio Del Moral Hernandez

138

Capacidade de reconhecimento de padrões em casos complexos NÃO LINEARES

Com as RNAs, a hipersuperfície de separação entre classes vai muito além dos hiperplanos e das hipercurvas mais simples

Em verde: fronteiras limitadas (implementadas com poucos nós neurais), que levam a alguns erros na classificação automática.



139

© Prof. Emilio Del Moral – EPUSP

139

(II) Mas ... Pode também haver taxa de acerto menor que 100% mesmo se tivermos possibilidade de fronteira com forma genérica:

- quando temos, no fenômeno que está sendo modelado, a sobreposição parcial das classes no espaço dos X;

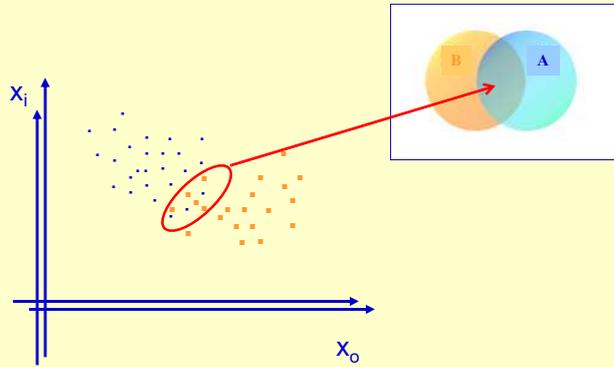
- quando isso ocorre, nenhuma fronteira de separação escolhida, por mais complexa que seja, garantirá a separação perfeita e a taxa de 100%.

140

© - Prof. Emilio Del Moral Hernandez

140

Sobreposição parcial de classes

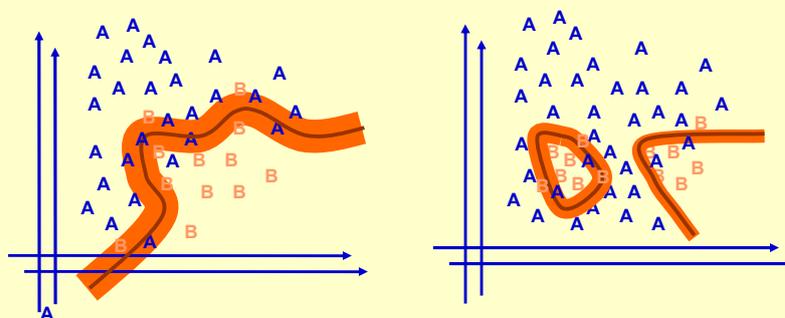


141

© - Prof. Emilio Del Moral Hernandez

141

Situações de classes com sobreposição parcial no espaço de atributos X ; situações de fronteiras de separação difusas ...



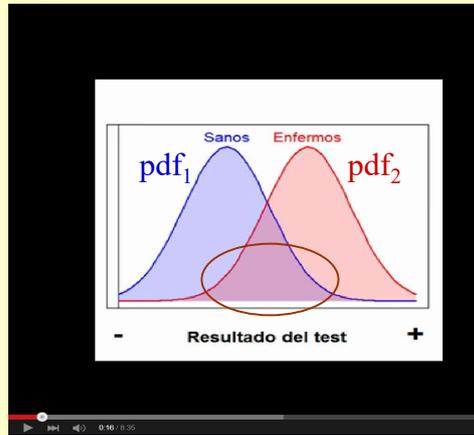
142

© Prof. Emilio Del Moral – EPUSP

142

Sobreposição parcial de classes

<https://www.youtube.com/watch?v=pA4E2uVHiYo> (início do vídeo)



Exemplo em diagnóstico médico:

Para certas faixas de valores, observamos que mesmo tendo-se exatamente o mesmo resultado clínico (tendo-se o mesmo valor de “nosso X”) pode ocorrer tanto de um indivíduo que obteve tal valor de resultado no exame estar SÃO (“nosso $y = -1$ ”) quanto de um outro com o mesmo valor de resultado estar ENFERMO (“nosso $y = +1$ ”). ... Podemos dizer que há sobreposição significativa das duas funções de densidade de probabilidade $f_{dp_1}(X|sadio)$ e $f_{dp_2}(X|enfermo)$.

144

© - Prof. Emilio Del Moral Hernandez

144

Matriz de confusão para caracterização de desempenho de reconhecedor de padrões

	“Classe A” (sugestão do MLP)	“Classe B” (sugestão do MLP)	
(Sabidamente) Classe A	AC_A	ER_A	✗
(Sabidamente) Classe B	ER_B	AC_B	✓

AC_A : padrão da classe A reconhecido como classe A

ER_A : padrão da classe A reconhecido como classe B

AC_B : padrão da classe B reconhecido como classe B

ER_B : padrão da classe B reconhecido como classe A

Taxa de Acertos Tente calculá-la com base nos elementos acima presentes na matriz de confusão

146

slide desenvolvido por Leandro Augusto da Silva

© - Prof. Emilio Del Moral Hernandez

146

Medidas de desempenho específicas para detectores

© - Prof. Emilio Del Moral Hernandez

149

Casos em que o classificador binário é encarado como um detector / identificador

- Uma das duas classes representa a detecção / o reconhecimento / a ocorrência de um dado cenário relevante e que está sendo alvo de detecção.
 - Por ex: detecção de provável doença num diagnóstico, detecção de uma potencial turbulência no mercado, detecção de uma provável fraude em curso, detecção de um potencial consumidor de um produto, etc etc etc.
- A outra classe, complementar à anterior, indica a **NÃO** ocorrência de tal cenário relevante e que está sendo alvo de detecção.
 - complementarmente: não há doença provável, não há potencial turbulência, não há provável fraude em curso, não é consumidor provável, etc etc etc.
- Nesse caso, uma classe é denominada positiva (= sim, temos detecção do cenário alvo) e a outra negativa (= não, o cenário alvo não foi detectado)

150

© - Prof. Emilio Del Moral Hernandez

150

Casos em que o classificador binário é encarado como um detector / identificador

- Assim, em lugar da saída binária do classificador automático ser encarada não mais como sendo a indicação de uma de duas classes abstratas A e B, mas sim como sendo a **indicação Positiva (P)** do detector ou a **indicação Negativa (N)** do detector, para a ocorrência de um alvo específico de interesse na detecção
- E temos expressões específicas associadas aos casos de **acertos** ou **erros** na classificação automática realizada, respectivamente **Classificações Verdadeiras (V)** e **Classificações Falsas (F)**

151

© - Prof. Emilio Del Moral Hernandez

151

Estendendo termos técnicos relacionados à matriz de confusão e suas componentes ...

- Dando destaque a uma das duas classes binárias: “Diagnóstico Positivo” / “Diagnóstico Negativo”
- Interpretações em outros contextos que não o médico:
 - Alarme de reconhecimento disparado x Alarme em silêncio
 - Risco provável previsto x Sem risco previsto
 - Necessidade de providência x Normalidade (sem necessidade)
 - Oportunidade de ação x Sem destaques que mereçam ação
- Contrastando o diagnóstico perfeito com o diagnóstico imperfeito de um exame ou de um autômato:
Diagnóstico Verdadeiro / Diagnóstico Falso
- Combinações dos diversos cenários VP / FP / VN / FN
 - Verdadeiro Positivo; Falso Positivo;
 - Verdadeiro Negativo; Falso Negativo.

153

© - Prof. Emilio Del Moral Hernandez

153

VP/FN/FP/VN ... Sensibilidade e Especificidade

<https://www.youtube.com/watch?v=y7Puf75vK0w>

The screenshot shows a YouTube video player with handwritten notes in blue and red ink. The notes are as follows:

Sensibilidade ? Especificidade ?

- **sensibilidade** : $\frac{VP}{D}$ → muito sensível : detecta bem os doentes **↑ FP**
90%.
- **especificidade** : $\frac{VN}{ND}$ → muito específico : detecta bem os não doentes **↓ FN**
90%.

Below the video player, the channel name "Canal Resumad" and the video title "Entenda o que é Sensibilidade e Especificidade #Dose Diária" are visible. The video has 5,554 views and 107 likes.

154

© - Prof. Emilio Del Moral Hernandez

154

VP/FN/FP/VN ... Sensibilidade e Especificidade

<https://www.youtube.com/watch?v=y7Puf75vK0w>

The screenshot shows the same YouTube video player as above, but with additional handwritten notes and a contingency table. The notes are as follows:

Sensibilidade ? Especificidade ?

- **sensibilidade** : $\frac{VP}{D}$ → muito sensível : detecta bem os doentes **↑ FP**
90%.
- **especificidade** : $\frac{VN}{ND}$ → muito específico : detecta bem os não doentes **↓ FN**
90%.

teste de detecção ⇒ GRIPE

A 2x2 contingency table is drawn in blue ink:

	doente	n-doente	
+	90 VP	10 FP	T
0	10 FN	90 VN	
	100 D	100 ND	200

Below the video player, the channel name "Canal Resumad" and the video title "Entenda o que é Sensibilidade e Especificidade #Dose Diária" are visible. The video has 5,554 views and 107 likes.

155

© - Prof. Emilio Del Moral Hernandez

155

VP/FN/FP/VN ... Sensibilidade e Especificidade (& curva ROC)

<https://www.youtube.com/watch?v=y7Puf75vK0w>

156

© - Prof. Emilio Del Moral Hernandez

156

Resumo ... Medidas de qualidade em reconhecedores de padrões / detectores

- Taxa de acertos (%)
- Taxa de erros (% complementar)
- Taxa de acertos em separado para as duas classes:
 - (A x B) ... Taxas de acerto e de erro para A e p / B
- Em situações em que uma das classes tem o significado de diagnóstico positivo e a outra de diagnóstico negativo, temos outras taxas consideradas relevantes ...
- Sensibilidade, Especificidade, etc etc
 - ... Sensibilidade = $VP / \text{Todos casos realmente Positivos}$
 - ... Especificidade = $VN / \text{Todos casos realmente Negativos}$

157

© - Prof. Emilio Del Moral Hernandez

157

Mais temas importantes ...

1 - O conceito de limiar ajustável em detectores

2 - A curva ROC associada ao exercício completo da gama de valores de limiar ajustável de um detector

166

© - Prof. Emilio Del Moral Hernandez

166

Estendendo conceitos relacionados à matriz de confusão e suas componentes VP, VN, FP, FN ...

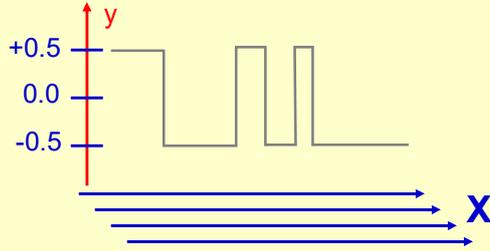
- **Conceito de limiar ajustável como parte do reconhecedor** (sensibilidade ajustável pelo usuário / variável de acordo com a circunstância)
- Indo do Reconhecedor / Detector super estressado – detector super sensível – ao Reconhecedor / Detector super dorminhoco – detector super específico – ...
(vamos de de 100% diagnósticos P a 100% diagnósticos N)
- Sensibilidade e Especificidade combinadas com o conceito de limiar ajustável leva ao conceito da curva ROC ...
Curva ROC de um reconhecedor com limiar ajustável:
Mapeamos a evolução dos valores de Sensibilidade e de Especificidade para todos os limiares, desde o extremo de detecção super sensível (baixo limiar) até o extremo oposto de detecção super específica (alto limiar), passando por todos os cenários intermediários de limiar.

167

© - Prof. Emilio Del Moral Hernandez

167

A função $f(X)$ a perseguir é desconhecida e assume valores binários



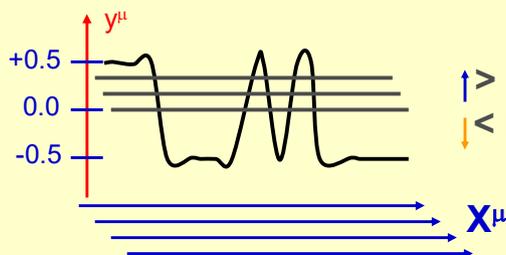
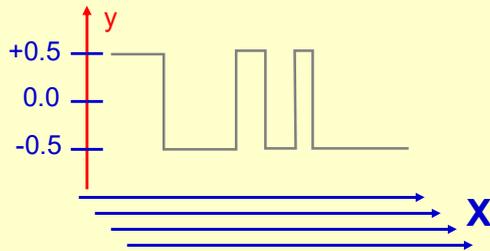
Sinal($y_{rede}(X)$)

168

© - Prof. Emilio Del Moral Hernandez

168

A função $f(X)$ a perseguir é desconhecida e assume valores binários



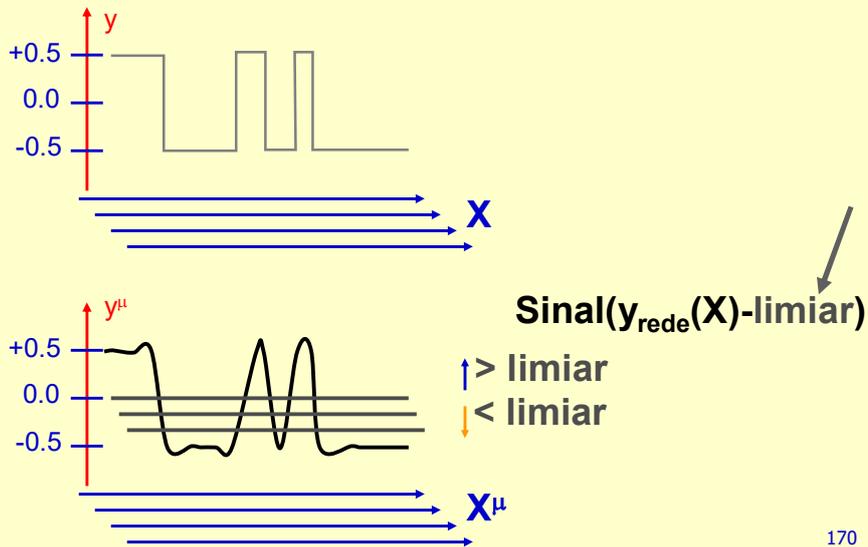
Sinal($y_{rede}(X)$ -limiar)

169

© - Prof. Emilio Del Moral Hernandez

169

A função $f(X)$ a perseguir é desconhecida e assume valores binários



170

© - Prof. Emilio Del Moral Hernandez

170

Podemos ter limiar variável em ...

- Redes neurais
- SVMs
- Comitês com composição (Adaboost e outros)
- Decisores estatísticos
- ...
- ...
- Sempre que houver uma grandeza de “saída” comparada com um limiar no estágio final, para a geração a classificação da decisão

Positivo versus Negativo

171

© - Prof. Emilio Del Moral Hernandez

171

Ilustrando o conceito de limiar ajustável em projetos concretos

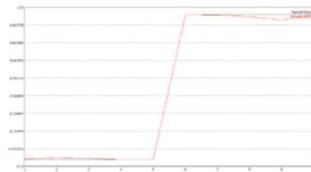
172

© - Prof. Emilio Del Moral Hernandez

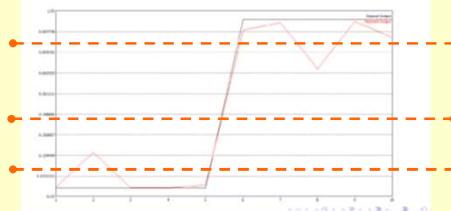
172

Carlos – Henrique – Pedro / detector de caracter (exemplo de limiar ajustável num projeto PSI2672 – práticas do 5º ano)

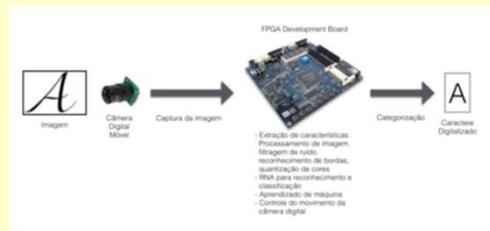
- Dimensão do vetor de entrada = 484 (uma entrada correspondente a cada elemento da matriz 22 x 22)
- Treinamento na RNA com 1 nó de função de transferência linear:



- Teste na RNA com 1 camada escondida formada por 10 nós de função de transferência tanh:



Conceito de limiar de detecção ajustável



174

© - Prof. Emilio Del Moral Hernandez

174

O conceito de curva ROC associada ao exercício completo do leque de valores do limiar ajustável em um detector

176

© - Prof. Emilio Del Moral Hernandez

176

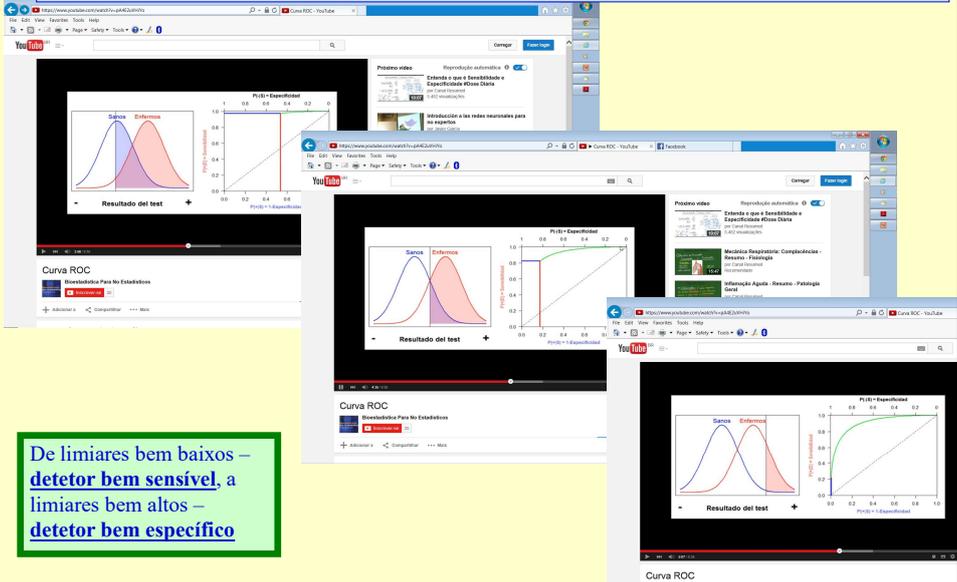
<https://www.youtube.com/watch?v=pA4E2uVHiYo> (início do vídeo)

Curva ROC

177

Conceito da curva ROC: construção ...

<https://www.youtube.com/watch?v=pA4E2uVHiYo> (início do vídeo)



The screenshot shows a YouTube video player with three overlapping windows. Each window displays a graph with two overlapping normal distributions (one blue, one red) and a corresponding ROC curve. The ROC curve plots True Positive Rate (TPR) on the y-axis against False Positive Rate (FPR) on the x-axis. A diagonal line represents a random classifier. The video player interface includes a title bar, navigation controls, and a video title 'Curva ROC'.

De limiares bem baixos – detetor bem sensível, a limiares bem altos – detetor bem específico

178

Area Under the Curve – AUC

A comparação geral de qualidades de detecção associadas as duas ROCs distintas é muitas vezes feita pelo contraste dos valores da área sob a curva ROC: AUC (Area Under the Curve)

Quanto maior a área sob a curva, melhor é a estratégia de detecção!! Através dessa medida, comparamos a qualidade de dois sistemas distintos que procuram detectar o mesmo objeto.

179

© - Prof. Emilio Del Moral Hernandez

179

Contexto: BCI. Duas ROCs contrastando pre-processamento distintos em um detetor de pulsos corticais (BCI) com limiar variável

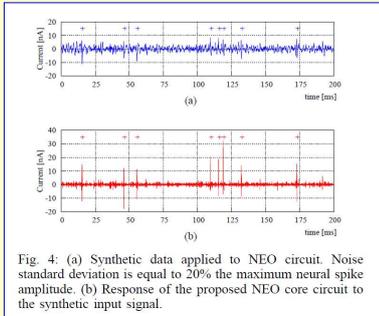


Fig. 4: (a) Synthetic data applied to NEO circuit. Noise standard deviation is equal to 20% the maximum neural spike amplitude. (b) Response of the proposed NEO core circuit to the synthetic input signal.

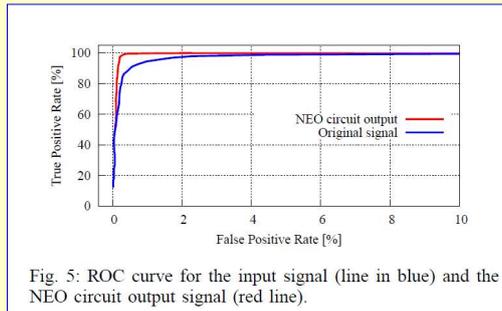


Fig. 5: ROC curve for the input signal (line in blue) and the NEO circuit output signal (red line).

*Trabalho do doutorando
Júlio Cesar Saldaña*

180

© - Prof. Emilio Del Moral Hernandez

180

Comparando diversas ROCs, de detectores distintos ...

<https://www.youtube.com/watch?v=pA4E2uVHiYo> (início do vídeo)

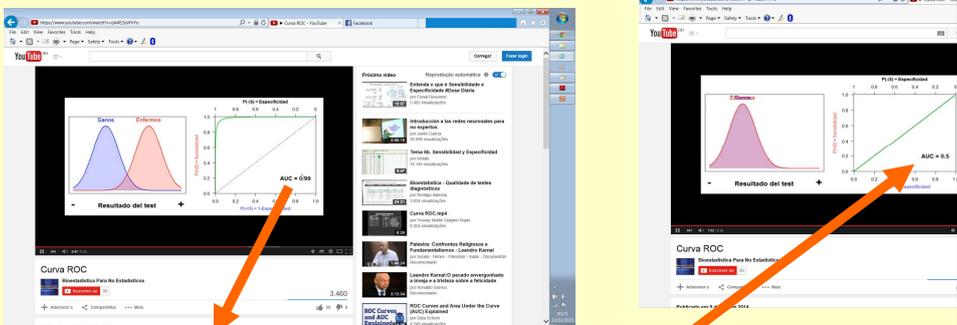
The collage shows three screenshots of a YouTube video. Each screenshot displays a ROC curve and a histogram. The first screenshot shows an AUC of 0.99. The second screenshot shows an AUC of 0.92. The third screenshot shows an AUC of 0.75. A green box with arrows pointing to the thumbnails contains the text: "As setas indicam qualidades cada vez mais pobres de discriminação".

© - Prof. Emilio Del Moral Hernandez

181

Comparando diversas ROCs, de detectores distintos ...

<https://www.youtube.com/watch?v=pA4E2uVHiYo> (início do vídeo)



AUC alta (0.99) – bom discriminador

AUC só 0.5 – o pior discriminador

182

© - Prof. Emilio Del Moral Hernandez

182

Importante lembrete:

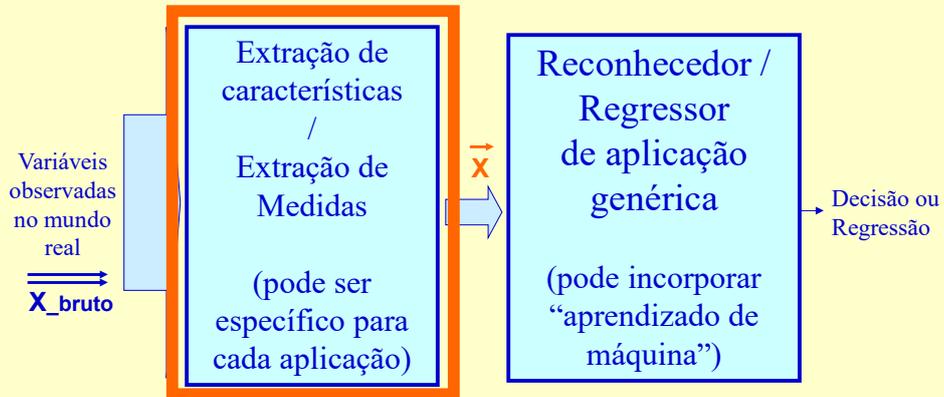
*diferentes estratégias de pré –
processamento e mesmo de
reconhecimento podem levar a
diferentes desempenhos de detecção
e portanto a diferentes ROCs*

183

© - Prof. Emilio Del Moral Hernandez

183

... O 1o estágio gera um Vetor de Medidas, \vec{X}
 (o segundo estágio operará sobre tal vetor)



184

© - Prof. Emilio Del Moral Hernandez

184

Exemplo concreto de efeito benéfico do pré-processamento no contexto de Brain Computer Interface, estudado pelo doutorando Júlio Cesar Saldaña

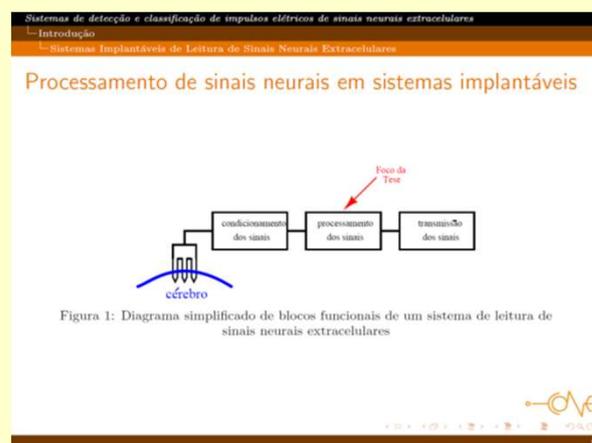


Figura 1: Diagrama simplificado de blocos funcionais de um sistema de leitura de sinais neurais extracelulares

185

© - Prof. Emilio Del Moral Hernandez

185

Contexto: BCI. Duas ROCs contrastando pre-processamento distintos em um detector de pulsos corticais (BCI) com limiar variável

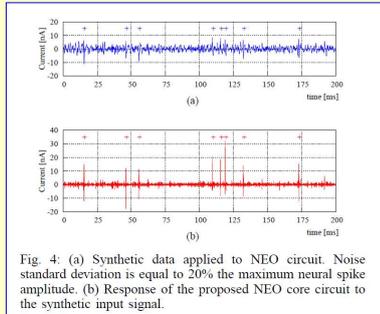


Fig. 4: (a) Synthetic data applied to NEO circuit. Noise standard deviation is equal to 20% the maximum neural spike amplitude. (b) Response of the proposed NEO core circuit to the synthetic input signal.

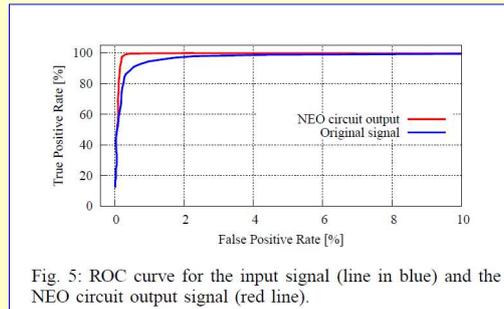


Fig. 5: ROC curve for the input signal (line in blue) and the NEO circuit output signal (red line).

*Trabalho do doutorando
Júlio Cesar Saldaña*

186

© - Prof. Emilio Del Moral Hernandez

186

Um parênteses final ...

Retornando à temática “Matriz de Confusão”, como você generalizaria o conceito de matriz de confusão para o caso de 3 classes em vez de apenas 2? E para 10 classes?

187

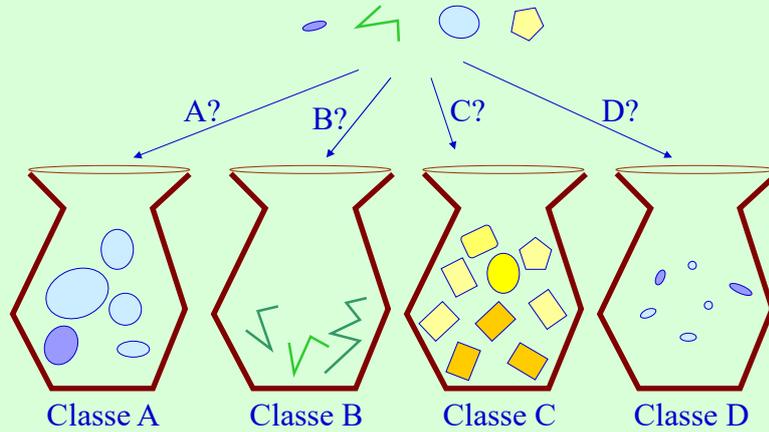
© - Prof. Emilio Del Moral Hernandez

187

Uma metáfora para a tarefa de classificação de padrões multiclasses (classificação não binária)...



Um dado objeto específico observado é de que tipo

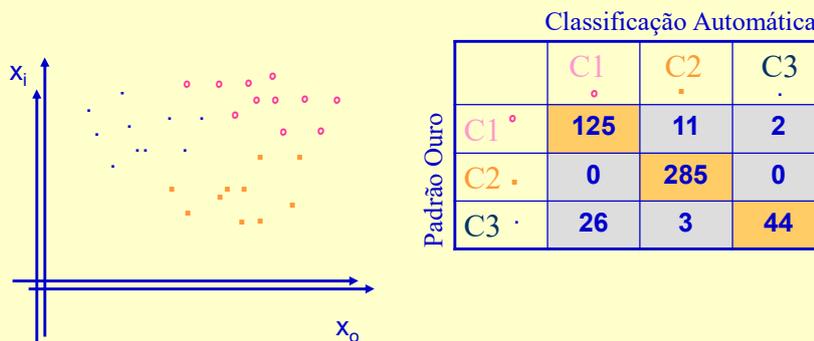


© Prof. Emilio Del Moral Hernandez

188

188

A matriz de confusão quando temos 3 (ou mais) classes



$$\text{Taxa de acertos global} = (125+285+44) / 495 = 92\%$$

(... aqui usamos a média simples – não ponderada – supondo o mesmo volume de exemplares empíricos nas três classes, C1, C2 e C3!)

189

© - Prof. Emilio Del Moral Hernandez

189

Destacando e contabilizando os volumes de confusões entre as diversas classes ...

- de C1 para C2: 11 (2%)
- de C1 para C3: 2 (0.4%)
- de C2 para C1: 0
- de C2 para C3: 0
- de C3 para C1: 26 (5%)
- de C3 para C2: 3 (0.6%)

Classificação Automática

	C1	C2	C3
Padrão Ouro C1	125	11	2
C2	0	285	0
C3	26	3	44

191

© - Prof. Emilio Del Moral Hernandez

191

193

Já finalizando ...

© Prof. Emilio Del Moral Hernandez

193

193

Emílio Del Moral Hernandez

*Nos falamos em breve ...
Prof. Emilio Del Moral Hernandez*

18:22 Role para ver detalhes

© Prof. Emilio Del Moral Hernandez

194

© Prof. Emilio Del Moral Hernandez

197