
Estatística: Aplicação ao Sensoriamento Remoto

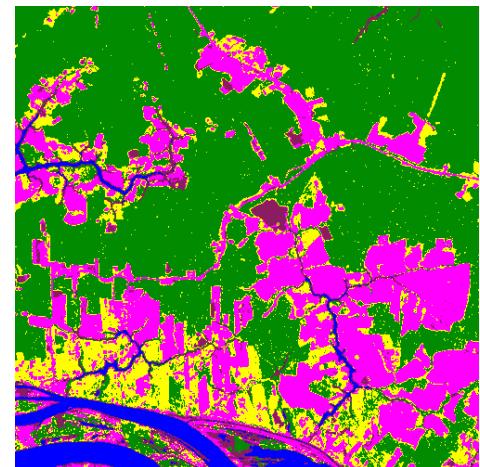
SER 204 - ANO 2024

Avaliação de Classificação

Camilo Daleles Rennó
camilo.renno@inpe.br
<http://www.dpi.inpe.br/~camilo/estatistica/>

Classificação de Imagens

Classificação é o processo de extração de informação em imagens para reconhecer padrões e objetos homogêneos que são utilizados para mapear áreas da superfície terrestre que correspondem aos temas de interesse.

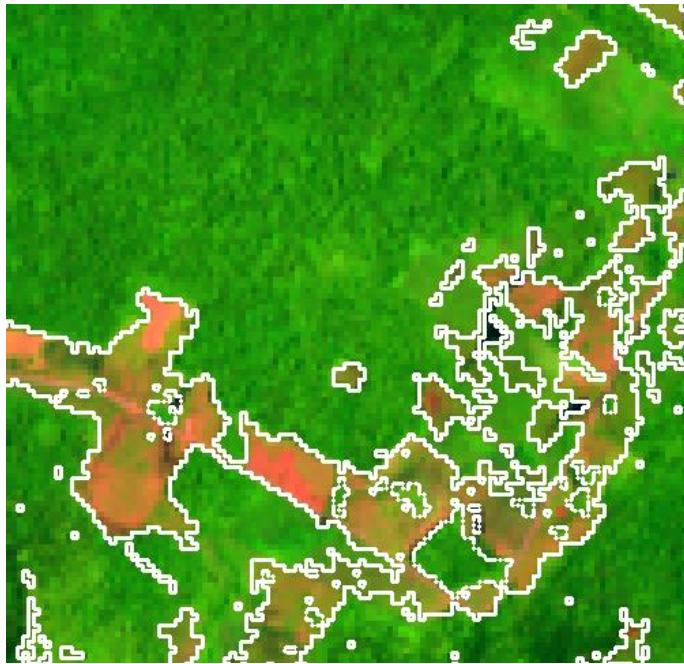


A partir de um conjunto de dados, através de um processo de rotulação (classificação), obtém-se uma **imagem classificada** (ou **mapa temático**).

Este processo inclui toda a preparação dos dados (seleção e pré-processamento), definição das classes de interesse, escolha do método de classificação e das funções que serão utilizadas pelo classificador para distinguir as classes, e a classificação propriamente dita (geração da imagem classificada).

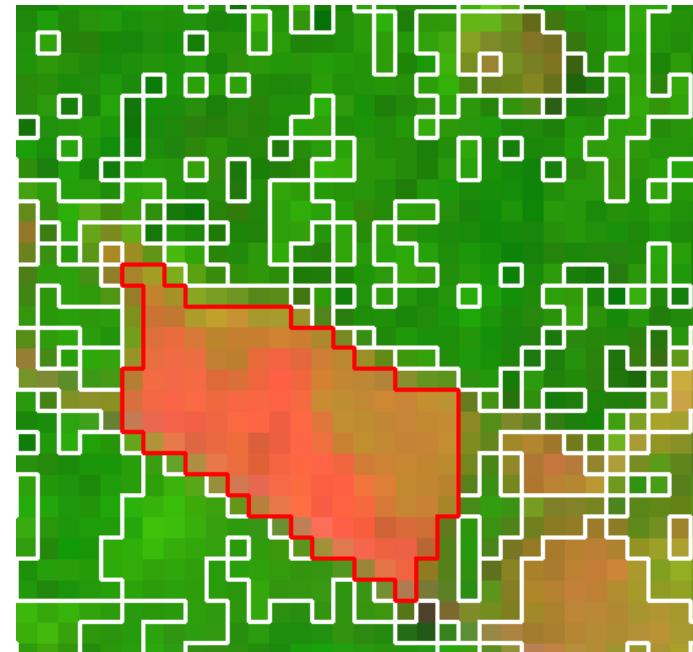
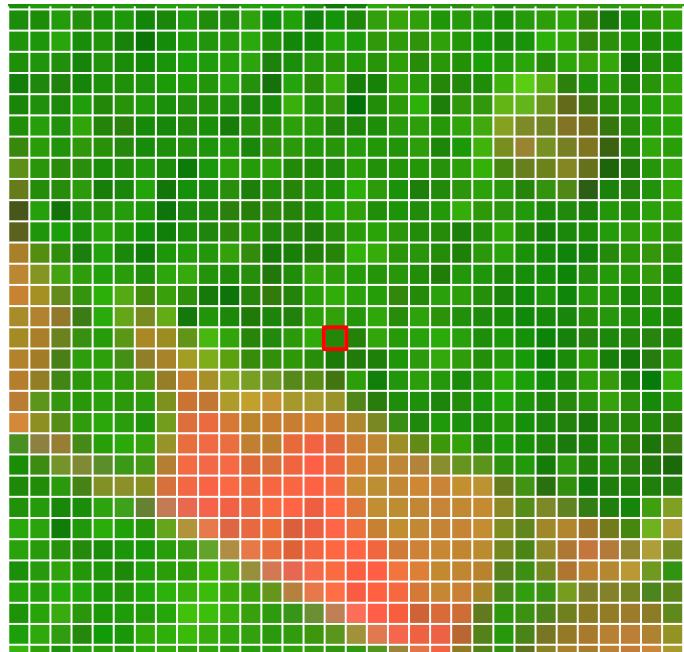
Tipos de Classificação

- Computacional ou interpretação visual



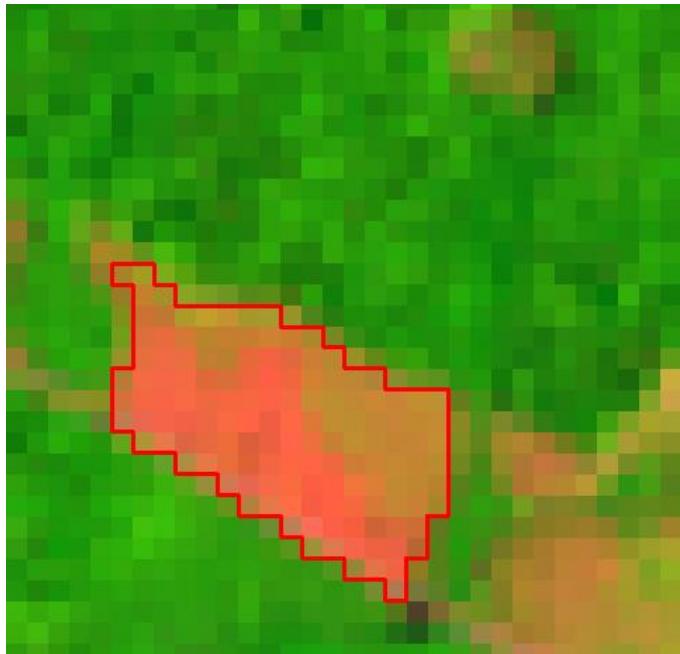
Tipos de Classificação

- Pixel a pixel ou por região



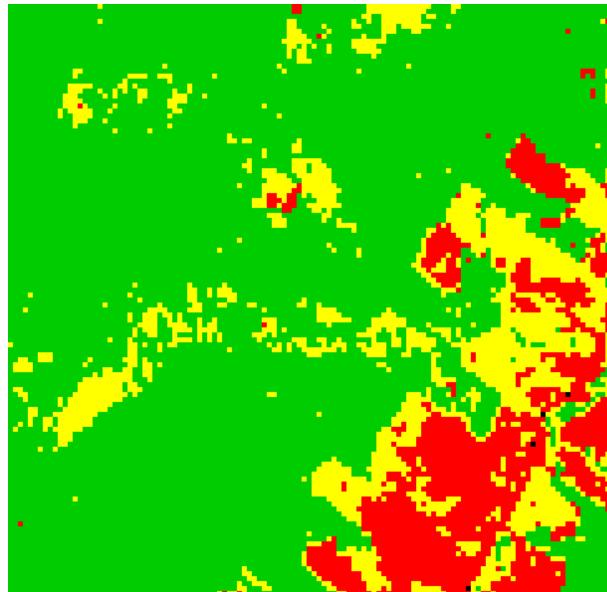
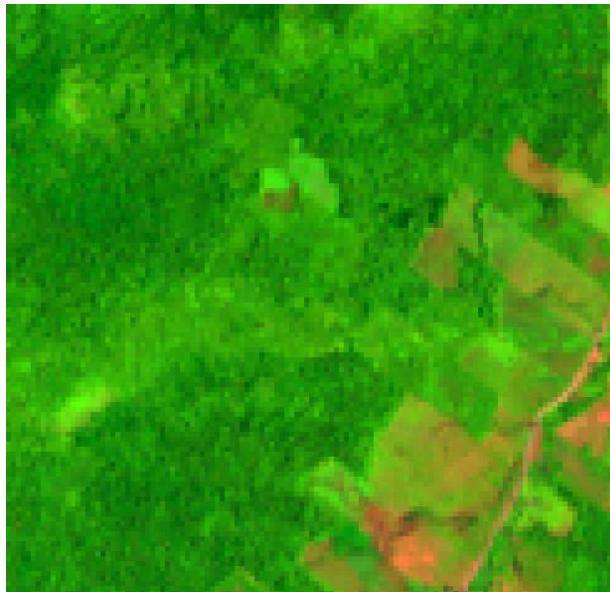
Tipos de Classificação

- Pontual ou contextual

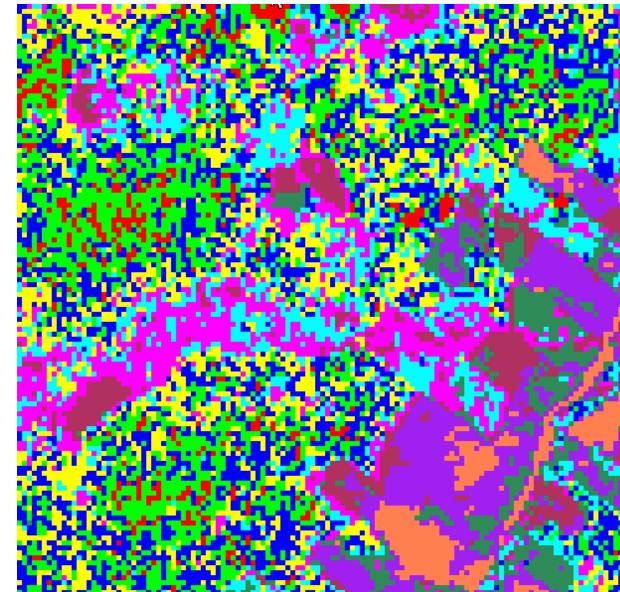


Tipos de Classificação

- Supervisionada ou automática (não supervisionada)



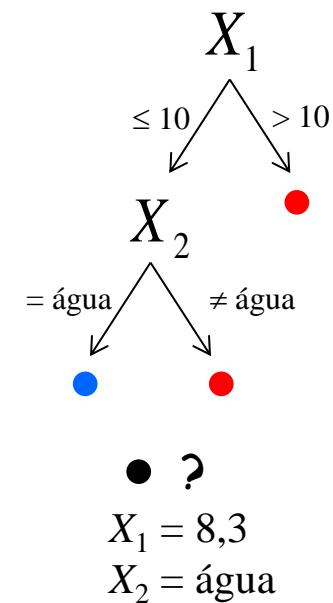
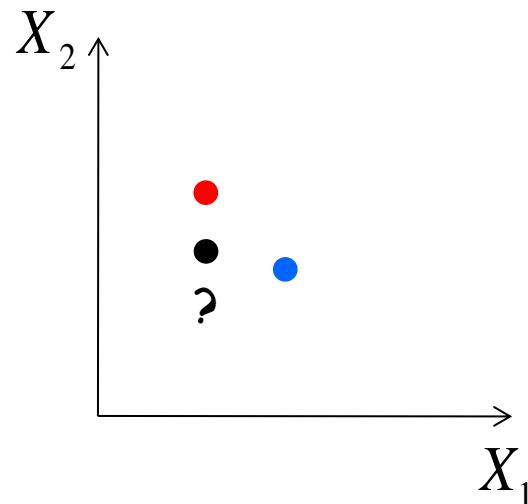
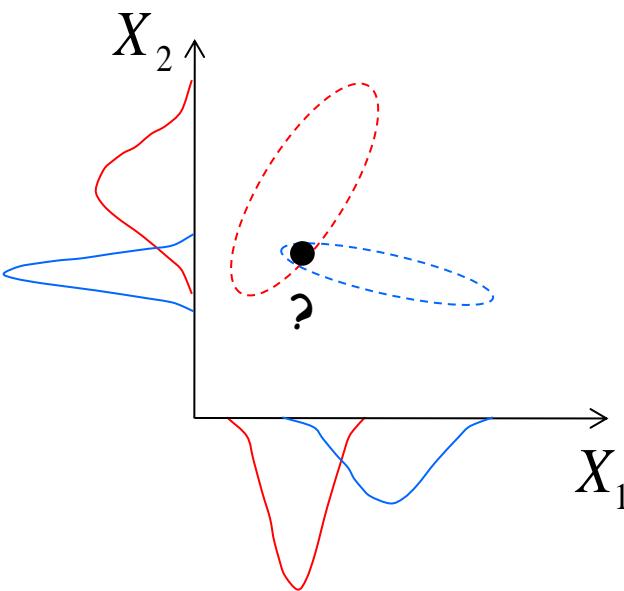
Supervisionado
Maxver Gaussiano (3 classes)



Não supervisionado
K-médias (10 classes)

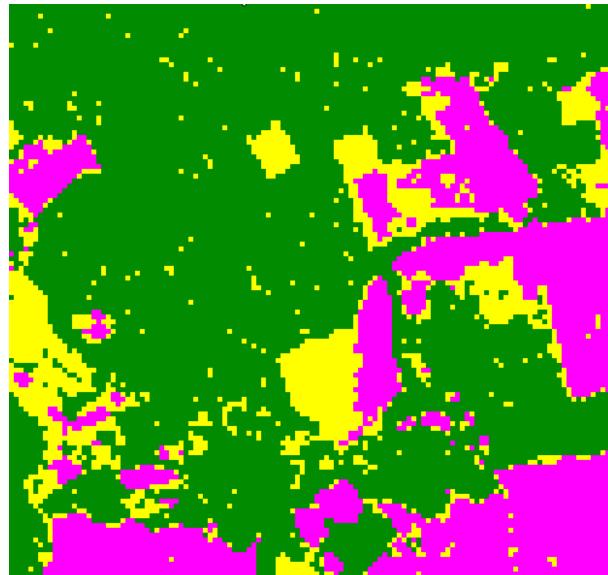
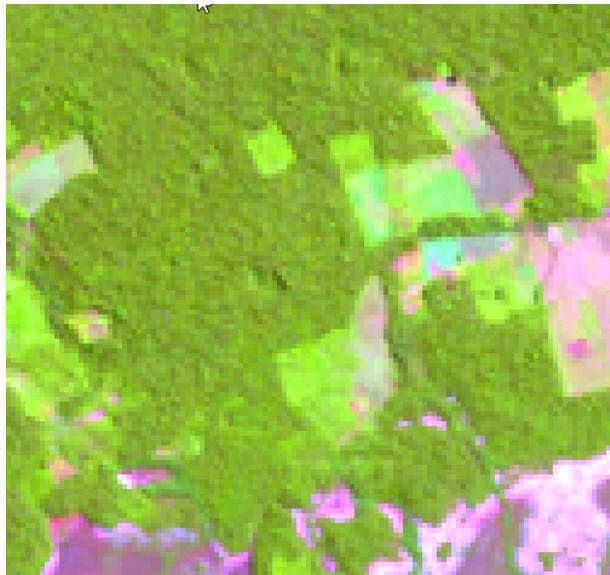
Tipos de Classificação

- Probabilística ou baseada em métricas de distância ou regras de decisão

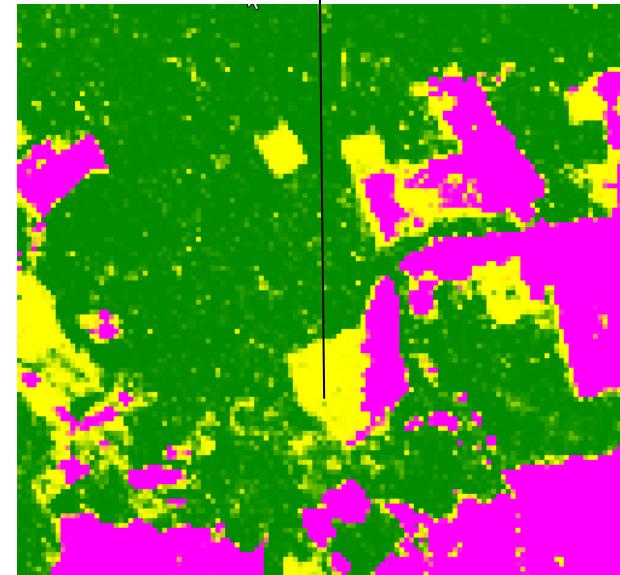


Tipos de Classificação

- Rígida (Discreta ou *Crisp*) ou Fuzzy (Contínua)



Maxver Gaussiano (3 classes)
Máxima Probabilidade



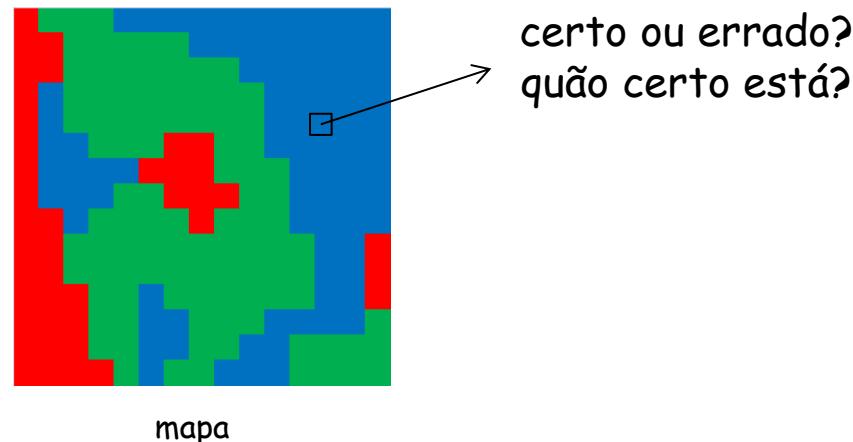
Maxver Gaussiano (3 classes)
Fuzzy - Probabilidade

Avaliação dos Erros de Classificação

Para classificadores supervisionados, as regras de decisão ou funções de pertinência são ajustadas segundo um conjunto de amostras de modo a minimizar **erros de classificação**

Numa classificação tradicional (rígida), considera-se que todo elemento (pixel ou polígono) está associado a uma única classe temática.

O erro surge sempre que esta associação diverge da **VERDADE**.



A **VERDADE**, muitas vezes, representa apenas uma **REFERÊNCIA** (resultado ideal).

Desafio

Como avaliar o desempenho de um classificador através do resultado da classificação, ou seja, através do mapa obtido?

O que estamos avaliando afinal?

Dados de entrada (atributos)?

dados são suficientes para descrever a complexidade da região estudada?

Método de classificação?

utiliza um método ou regras que fazem sentido para os meus dados?

Poder discriminatório do classificador?

a definição das classes é realmente clara?

Qualidade das amostras fornecidas?

a seleção das amostras foi bem feita? São representativas? São realmente confiáveis?

Questões básicas

- Amostragem

O que amostrar?

Quanto amostrar?

Como amostrar?

Como separar as amostras entre as diferentes fases de classificação/avaliação?

- Índices de avaliação

Como representar erros e acertos?

Como representar incertezas?

Melhor usar índices globais ou por classe?

Há índices apropriados para avaliar uma classificação Fuzzy?

Como se pode comparar classificadores?

- O que não abordaremos?

Acurácia posicional

Especificação Técnica para Controle de Qualidade de Dados Geoespaciais

Erro e Incerteza

Na classificação, a incerteza surge devido a diferentes fatores: resolução espacial (mistura de classes), definição das classes, escolha de amostras representativas, escolha de atributos representativos (capazes de discriminar as classes de interesse), escolha das funções discriminantes ou de pertinência (e de seus parâmetros), erros aleatórios e sistemáticos, etc.

Existem vários tipos de incerteza e o significado atribuído ao termo incerteza aparece com diferentes interpretações na literatura: **erro**, **ambiguidade** ou **imprecisão**.

Avaliação dos Erros de Classificação

avalia os erros e confusões entre classes de acordo com uma referência
Exatidão ou Acurácia

Avaliação da Incerteza

avalia como as questões estocásticas afetam o resultado da classificação
Mapas de Incerteza

Avaliação dos Erros de Classificação

A referência pode ser obtida a partir de:

- dados pré-existentes (levantamentos, mapas, literatura, etc)
apesar de ter custo quase zero, as informações podem estar desatualizadas
erros pré-existentes são raramente conhecidos
pode haver diferentes definições para a mesma classe (diferença semântica)
- dados de campo
em geral, envolve um custo elevado (logística, localização precisa, equipes grandes, etc)
pode haver grande defasagem temporal entre o momento de obtenção do dado usado
na classificação e o de checagem (comum em ambientes dinâmicos)
a amostragem pode ser enviesada
somente pontos com fácil acesso são checados
o uso de drones pode ajudar nesse caso
- fotointerpretação (geralmente usando imagens com resolução mais fina)
apesar do baixo custo, o resultado depende da experiência do fotointérprete (ideal:
diferente de quem fez a classificação)
pode ser enviesado ao induzir um resultado positivo quando já se conhece o resultado
da classificação (ideal: total independência entre os processos)
também pode ter problemas em ambientes dinâmicos (defasagem temporal)

Avaliação dos Erros de Classificação

Fatores importantes a serem considerados durante a avaliação dos erros:

- Unidade amostral utilizada na avaliação
pontos, objetos (polígonos ou linhas), grupo de pontos ou objetos
importante: tem impacto direto sobre a interpretação dos resultados
- Independência das amostras
supõe-se que todos pontos (ou objetos) tenham a mesma probabilidade (não nula) de serem sorteados
a autocorrelação espacial tem impacto maior se for utilizar testes estatísticos de significância pois afeta principalmente o cálculo da precisão (variância) associada aos índices testados
- Representação dos acertos/erros
como explicitar os erros e acertos? Índices globais, por classe e/ou distribuição espacial?

Amostragem - recomendações

Tamanho da amostra

- tamanhos ideais de amostras podem ser calculados com base na distribuição binomial, considerando-se os erros do tipo I e/ou II. Essas abordagens desconsideram as confusões entre classes;
- Congalton (1991)* e Olofsson et al. (2014)** sugerem no mínimo 50 pontos para cada classe num caso geral, e 100 pontos caso a área avaliada seja muito grande ou o número de classes seja maior do que 12.

Tipo de amostragem

- há relação direta entre o tipo de amostragem e a interpretação dos resultados;
- Congalton (1991)* sugere as amostragens aleatória simples e estratificada como as mais promissoras.

*Congalton, R. G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.*, 37(1):35-46, 1991

**Olofsson, P. et al. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* 148:42-57, 2014

Representação dos Acertos/Erros

A avaliação dos acertos/erros de uma classificação pode ser feita através de:

- Tabelas (matriz de confusão) - representa os acertos e confusões/erros entre classes
- Índices globais - representam características gerais
- Índices por classes - representam características particulares de cada classe
- Índice F1 score* - representa o balanço entre precisão e sensibilidade (*recall*)
- Curva ROC* - representa relação entre falsos e verdadeiros positivos

* específico para classificações com 2 classes (presente/ausente)

Matriz de Confusão

A partir de um conjunto de pontos (polígonos ou grupos) sorteados/escolhidos, compara-se a classificação e a referência e constrói-se a

Matriz de Confusão (matriz de erro)

		Referência				
		C ₁	C ₂	...	C _c	Total
Classificação	C ₁	x ₁₁	x ₁₂	...	x _{1c}	x ₁₊
	C ₂	x ₂₁	x ₂₂	...	x _{2c}	x ₂₊
	:	:	:	⋮	⋮	⋮
	C _c	x _{c1}	x _{c2}	...	x _{cc}	x _{c+}
	Total	x ₊₁	x ₊₂	...	x _{+c}	n

x_{ij} : número de pontos da classe j (referência), classificados na classe i (classificação)

x_{kk} : número total de pontos corretamente classificados da classe k

x_{+j} : número total de pontos avaliados da classe j na referência

x_{i+} : número total de pontos avaliados da classe i na classificação

Matriz de Confusão

Observações importantes:

- considera que as classes são excludentes (cada ponto pertence a apenas uma classe);
- mesmas classes na Classificação e na Referência (matriz quadrada);
- todos os pontos avaliados devem pertencer a alguma classe, ou seja, o classificador não pode considerar a classe “não classificado”;
- a utilização de classes muito semelhantes (representando subtipos) pode induzir a um excesso de erros (ou confusões) que certamente prejudicarão a avaliação global da classificação;
- a interpretação dos resultados está diretamente dependente da unidade amostral adotada (pontos ou polígonos isolados ou em grupos).

Avaliação da Exatidão

		Referência					
		A	B	C	D	Total	
Classificação	A	13	8	0	21	21	#pontos corretamente classificados
	B	8	10	0	3	21	#pontos/classe avaliados na classificação
	C	0	5	27	4	36	
	D	0	0	0	0	0	#pontos incorretamente classificados
	Total	21	23	27			#pontos amostrados

#pontos/classe avaliados na referência

Avaliação da Exatidão

		Referência			
		A	B	C	D
Classificação	A	13	8	0	0
	B	8	10	0	3
	C	0	5	27	4
	D	0	0	0	32
	Total	21	23	27	39
		Total			

$$\text{Exatidão Total} = \frac{\sum_{k=1}^c x_{kk}}{n} \quad \left\{ \begin{array}{l} \text{mínimo} = 0 \\ \text{máximo} = 1 \text{ (ou 100\%)} \end{array} \right.$$

$$= \frac{13+10+27+32}{110} = \frac{82}{110} = 74,5\%$$

Avaliação da Exatidão

		Referência				
		A	B	C	D	Total
Classificação	A	13	8	0	0	21
	B	8	10	0	3	21
	C	0	5	27	4	36
	D	0	0	0	32	32
	Total	21	23	27	39	110

$$\text{Exatidão} = \frac{\sum_{k=1}^c x_{kk}}{n} = 74,5\%$$

O que significa uma exatidão de 74,5% ?

- Se uma amostra, dentre as 110, fosse escolhida ao acaso, a probabilidade desta estar corretamente classificada seria de 74,5%
- Se uma amostra (ponto ou polígono) fosse escolhida ao acaso no mapa, a probabilidade desta estar corretamente classificada seria também de 74,5%?
somente se a amostragem representar as reais proporções de cada classe!

Avaliação da Exatidão

		Referência			
		A	B	C	D
Classificação	A	13	8	0	0
	B	8	10	0	3
	C	0	5	27	4
	D	0	0	0	32
	Total	21	23	27	39
		Total			

Pontius e Millones (2011) sugerem sempre utilizar a matriz não enviesada (ajustada para as **proporções reais** de cada classe)

	Proporção			Exatidão 74,5%
	A	B	C	
A	25%	1%	1%	
B	25%	2%	95%	
C	25%	95%	3%	
D	25%	2%	1%	
exatidão	71,9%	98,1%	45,7%	

Problema: proporções reais são quase sempre desconhecidas!

Avaliação da Exatidão por Classe

Ponto de vista do Produtor (Referência)

Quanto da classe k foi identificada corretamente pelo classificador?

		Referência				
		A	B	C	D	Total
Classificação	A	13	8	0	0	21
	B	8	10	0	3	21
	C	0	5	27	4	36
	D	0	0	0	32	32
	Total	21	23	27	39	110

$$\text{Exatidão do Produtor da classe } k = \frac{x_{kk}}{x_{+k}}$$

$$\text{Exatidão do Produtor da classe B} = \frac{10}{23} = 43,5\%$$

$$\text{Erro de omissão da classe B} = \frac{8+5+0}{23} = \frac{13}{23} = 56,5\%$$

Avaliação da Exatidão por Classe

		Referência				
		A	B	C	D	Total
Classificação	A	13	8	0	0	21
	B	8	10	0	3	21
	C	0	5	27	4	36
	D	0	0	0	32	32
	Total	21	23	27	39	110

Ponto de vista do Usuário (Classificação)

Quanto do que foi classificado como k é realmente da classe k ?

$$\text{Exatidão do Usuário da classe } k = \frac{x_{kk}}{x_{k+}}$$

$$\text{Exatidão do Usuário da classe B} = \frac{10}{21} = 47,6\%$$

$$\text{Erro de comissão (inclusão) da classe B} = \frac{8+0+3}{21} = \frac{11}{21} = 52,4\%$$

Avaliação da Exatidão por Classe

não é necessário apresentar a exatidão e o erro simultaneamente!

		Referência				Classificação	
		A	B	C	D	Total	
A	A	13	8	0	0	21	
	B	8	10	0	3	21	
C	C	0	5	27	4	36	
	D	0	0	0	32	32	
Total		21	23	27	39	110	

Classe	exatidão	erro	exatidão	erro
	produtor	omissão	usuário	inclusão
A	61,9%	38,1%	61,9%	38,1%
B	43,5%	56,5%	47,6%	52,4%
C	100,0%	0,0%	75,0%	25,0%
D	82,1%	17,9%	100,0%	0,0%

Classe C:

tudo que é realmente C está no mapa, mas nem tudo que está no mapa como C é realmente C

Classe D:

tudo que está no mapa como D é realmente D, mas nem tudo que é realmente D está no mapa

Classe A:

erros de omissão e inclusão se compensam.
Apesar dos erros, o mapa estima bem a área ocupada por essa classe!

Avaliação da Exatidão

		Referência			
		A	B	C	D
Classificação	A	13	8	0	0
	B	8	10	0	3
	C	0	5	27	4
	D	0	0	0	32
	Total	21	23	27	39
		Total			

$$\text{Exatidão Total} = 74,5\%$$

E se a classificação fosse realizada de modo totalmente aleatório?

$$P(C_i \cap R_j) = P(C_i).P(R_j)$$

Avaliação da Exatidão

$$\frac{21*21}{110}$$

		Referência			
		A	B	C	D
Classificação	A	4,01	4,39	5,15	7,45
	B	4,01	4,39	5,15	7,45
	C	6,87	7,53	8,84	12,76
	D	6,11	6,69	7,85	11,35
	Total	21	23	27	39
		Total			

$$\text{Exatidão Total} = \frac{4,01 + 4,39 + 8,84 + 11,35}{110} = \frac{28,59}{110} = 26\%$$

Ou seja, 26% do acerto pode ter sido conseguido de modo casual !!!

Medida de Concordância Kappa

		Referência				
		A	B	C	D	Total
Classificação	A	13	8	0	0	21
	B	8	10	0	3	21
	C	0	5	27	4	36
	D	0	0	0	32	32
	Total	21	23	27	39	110

Índice Kappa (κ) - medida de concordância

$$\hat{\kappa} = \frac{\theta_1 - \theta_2}{1 - \theta_2} \quad \theta_1 = \frac{\sum_{k=1}^c x_{kk}}{n}$$

exatidão total
(observada)

$$\kappa \begin{cases} \text{mínimo} = < 0 & (\theta_1 < \theta_2) \\ \text{máximo} = 1 & (\theta_1 = 1) \end{cases} \quad \theta_2 = \frac{\sum_{k=1}^c x_{k+} x_{+k}}{n^2}$$

exatidão total
(se classificação fosse aleatória)

Índice Kappa

		Referência				
		A	B	C	D	Total
Classificação	A	13	8	0	0	21
	B	8	10	0	3	21
	C	0	5	27	4	36
	D	0	0	0	32	32
	Total	21	23	27	39	110

Índice Kappa (κ) - medida de concordância

$$\theta_1 = \frac{\sum_{k=1}^c x_{kk}}{n} = 0,745 \quad \hat{\kappa} = \frac{\theta_1 - \theta_2}{1 - \theta_2} = \frac{0,745 - 0,260}{1 - 0,260} = 0,6561$$

$$\theta_2 = \frac{\sum_{k=1}^c x_{k+} x_{+k}}{n^2} = 0,260$$

Se a classificação fosse totalmente aleatória, qual seria o valor esperado para o kappa? zero \Rightarrow Teste de hipótese

Índice Kappa

$$\hat{\kappa} = \frac{\theta_1 - \theta_2}{1 - \theta_2} \quad \theta_1 = \frac{\sum_{k=1}^c x_{kk}}{n} \quad \theta_2 = \frac{\sum_{k=1}^c x_{k+} x_{+k}}{n^2}$$

$$Var(\hat{\kappa}) = \frac{1}{n} \left[\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2^2)}{(1-\theta_2)^4} \right]$$

$$\theta_3 = \sum_{k=1}^c x_{kk} (x_{k+} + x_{+k}) / n^2$$

$$\theta_4 = \sum_{i=1}^c \sum_{j=1}^c x_{ij} (x_{i+} + x_{+j})^2 / n^3$$

Pressupondo **amostras independentes e TLC válido** (amostra grande):

$$Z = \frac{\hat{\kappa} - \kappa}{\sqrt{Var(\hat{\kappa})}} \sim N(0,1)$$

Índice Kappa - Exemplo 1

A fim de avaliar a qualidade de uma classificação, 150 pontos foram selecionados aleatoriamente, avaliando-se cada ponto segundo uma referência. O resultado é apresentado na matriz de confusão abaixo:

		referência				
		A	B	C	D	E
classificação	A	13	0	3	0	0
	B	8	10	5	0	0
	C	8	4	27	0	0
	D	2	0	1	25	0
	E	0	0	0	0	44

Verifique se há mesmo uma concordância entre a classificação e a referência através de um teste de hipótese para o Kappa.

Índice Kappa - Exemplo 1

		referência				
		A	B	C	D	E
classificação	A	13	0	3	0	0
	B	8	10	5	0	0
	C	8	4	27	0	0
	D	2	0	1	25	0
	E	0	0	0	0	44

$$\hat{\kappa} = 0,7364$$

$$\text{Var}(\hat{\kappa}) = 0,001664$$

$$Z = \frac{\hat{\kappa} - \kappa}{\sqrt{\text{Var}(\hat{\kappa})}} \sim N(0,1)$$

$$H_0: \kappa = 0$$

$$H_1: \kappa > 0$$

Se H_0 for verdadeira, então:

$$z = \frac{\hat{\kappa}}{\sqrt{\text{Var}(\hat{\kappa})}} = \frac{0,7364}{\sqrt{0,001664}} = 18,05 \quad \text{valor-P} = 0,0000$$

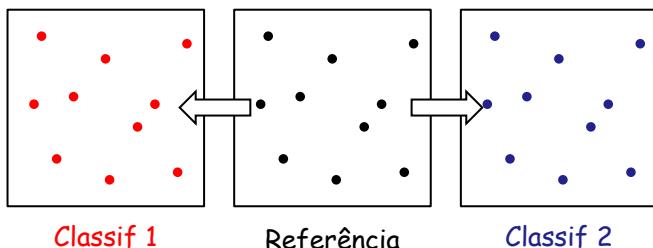
Conclusão: rejeita-se H_0 com 5% de significância, ou seja, há concordância entre a classificação e a referência

Comparando Avaliações...

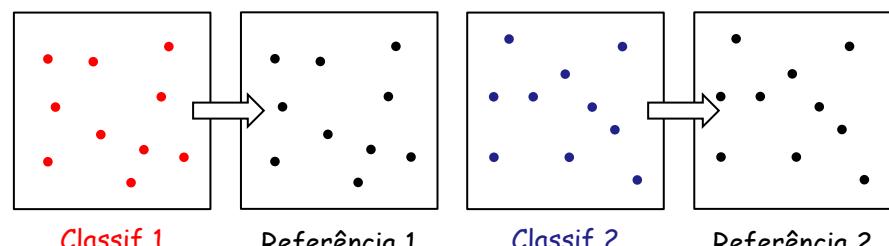
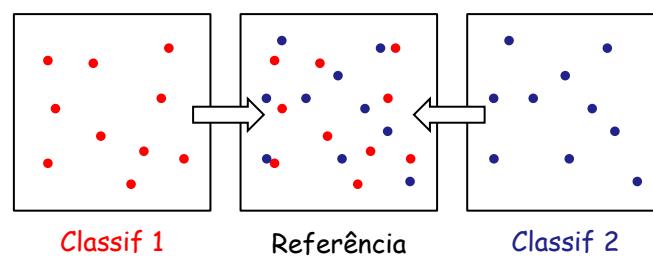
Em muitos casos, deseja-se verificar se duas classificações apresentam o mesmo desempenho ou se uma das classificações tem resultados mais favoráveis que outra.

Uma questão fundamental aqui é saber se as amostras utilizadas para as avaliações são as mesmas ou cada classificação foi avaliada com um conjunto diferente de amostras, ou seja, as duas classificações foram avaliadas de forma independente uma da outra.

Avaliação usando mesmas amostras



Avaliação usando amostras independentes



Avaliação a partir de 2 amostragens não relacionadas

Nesse caso, as avaliações são **independentes** entre si!

Pressupondo-se **TLC válido** (amostra grande), tem-se

- para a exatidão:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0,1) \quad \hat{p} = \frac{\hat{p}_1 + \hat{p}_2}{2}$$

- para o Kappa:

$$Z = \frac{(\hat{\kappa}_1 - \hat{\kappa}_2) - (\kappa_1 - \kappa_2)}{\sqrt{Var(\hat{\kappa}_1) + Var(\hat{\kappa}_2)}} \sim N(0,1)$$

Avaliação a partir de 2 amostragens relacionadas

Nesse caso, as avaliações foram feitas usando-se as mesmas amostras de referência para ambas classificações

- para a exatidão:

teste de McNemar (Foody, 2004)

		Classificação 2	
		Correto	Incorreto
Classificação 1	Correto	f_{11}	f_{12}
	Incorreto	f_{21}	f_{22}

$$\chi^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \sim \chi_1^2$$

Teste qui-quadrado unilateral a direita

- para a Kappa:

Não há um teste formal para comparar dois índices Kappa para amostras relacionadas!

Índice Kappa - Exemplo 2

Dois classificadores foram testados. Há uma suposição de que o segundo classificador gera resultados superiores (melhor exatidão) quando comparado ao primeiro classificador. Para checar esta hipótese, 150 pontos foram escolhidos aleatoriamente para cada avaliação. As matrizes de confusão resultantes das avaliações são apresentadas a seguir:

		referência				
		A	B	C	D	E
classificação 1	A	13	0	3	0	0
	B	8	10	5	0	0
	C	8	4	27	0	0
	D	2	0	1	25	0
	E	0	0	0	0	44

		referência				
		A	B	C	D	E
classificação 2	A	25	0	2	0	0
	B	2	28	0	0	4
	C	0	2	30	2	0
	D	0	0	1	33	0
	E	0	0	0	0	21

Verifique se o segundo classificador realmente apresenta uma concordância com a referência superior a do primeiro classificador comparando-se os valores Kappa de cada avaliação.

Índice Kappa - Exemplo 2

		referência				
		A	B	C	D	E
classificação 1	A	13	0	3	0	0
	B	8	10	5	0	0
	C	8	4	27	0	0
	D	2	0	1	25	0
	E	0	0	0	0	44

$$\hat{\kappa}_1 = 0,7364 \quad \text{Var}(\hat{\kappa}_1) = 0,001664$$

$$Z = \frac{(\hat{\kappa}_1 - \hat{\kappa}_2) - (\kappa_1 - \kappa_2)}{\sqrt{\text{Var}(\hat{\kappa}_1) + \text{Var}(\hat{\kappa}_2)}} \sim N(0,1)$$

$$H_0: \kappa_1 - \kappa_2 = 0$$

$$H_1: \kappa_1 - \kappa_2 < 0$$

Se H_0 for verdadeira, então:

$$z = \frac{\hat{\kappa}_1 - \hat{\kappa}_2}{\sqrt{\text{Var}(\hat{\kappa}_1) + \text{Var}(\hat{\kappa}_2)}} = \frac{0,7364 - 0,8911}{\sqrt{0,001664 + 0,000831}} = -3,10 \quad \text{valor-P} = 0,0010$$

Conclusão: rejeita-se H_0 com 5% de significância, ou seja, a concordância entre a classificação e a referência é maior para o classificador 2

		referência				
		A	B	C	D	E
classificação 2	A	25	0	2	0	0
	B	2	28	0	0	4
	C	0	2	30	2	0
	D	0	0	1	33	0
	E	0	0	0	0	21

$$\hat{\kappa}_2 = 0,8911 \quad \text{Var}(\hat{\kappa}_2) = 0,000831$$

Importante: amostragens devem ser independentes.
Ou seja, cada avaliação deve usar um conjunto diferente de validação!!!

Independência das Amostras

Outra questão importante diz respeito a independência (espacial e temporal) na coleta das amostras, pois amostras não independentes representam redundância e portanto podem gerar problemas principalmente na estimativa das incertezas (variância) associadas ao índice avaliado.

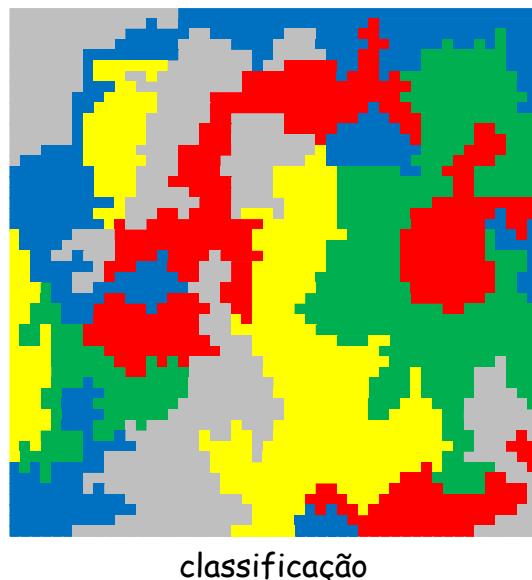
Em geral, um grande número de amostras correlacionadas geram a falsa impressão de que as estimativas são muito precisas!

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

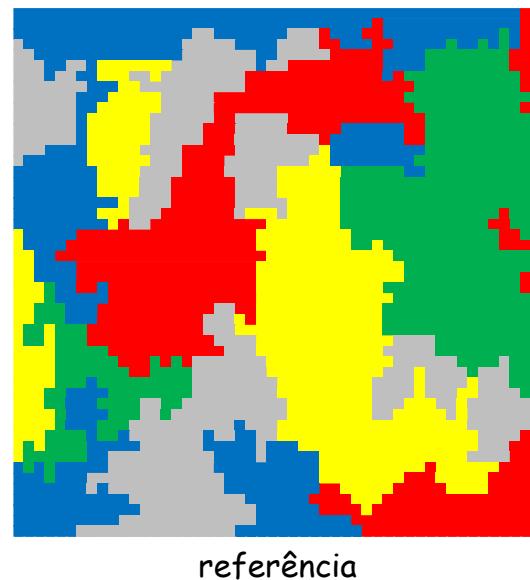
$$Var(\hat{\kappa}) = \frac{1}{n} \left[\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2^2)}{(1-\theta_2)^4} \right]$$

Índice Kappa - Exemplo 3

Avalie se a classificação abaixo apresenta Kappa superior a 0,7, considerando-se a existência de uma referência que recobre toda a área classificada.

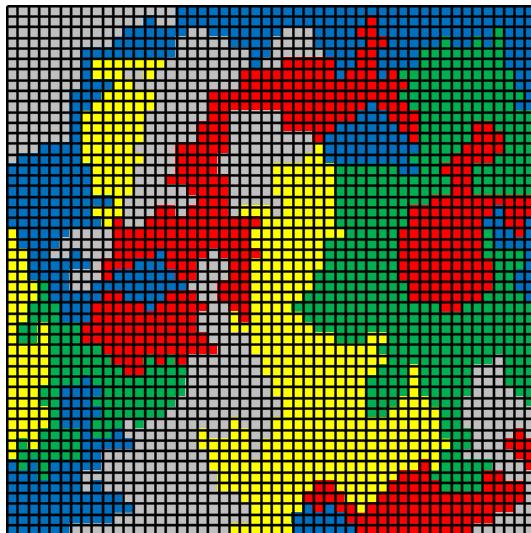


classificação

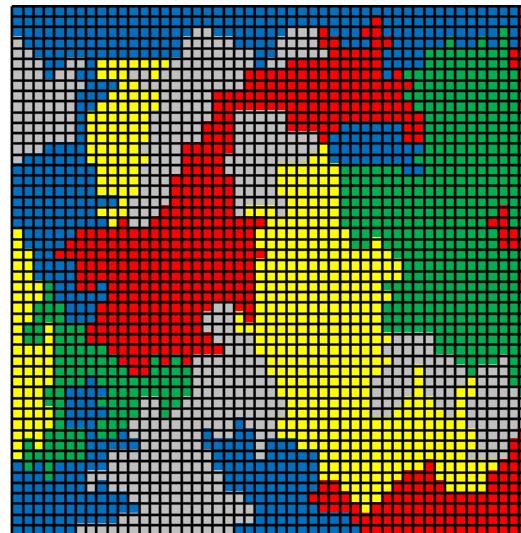


referência

Índice Kappa - Exemplo 3



classificação



referência

Selecionando-se
todos os pontos...

$$H_0: \kappa = 0,7$$

$$H_1: \kappa > 0,7$$

$$\hat{\kappa} = 0,7400$$

$$\text{Var}(\hat{\kappa}) = 0,000103$$

Se H_0 for verdadeira, então:

$$z = \frac{\hat{\kappa} - 0,7}{\sqrt{\text{Var}(\hat{\kappa})}} = 3,9475$$

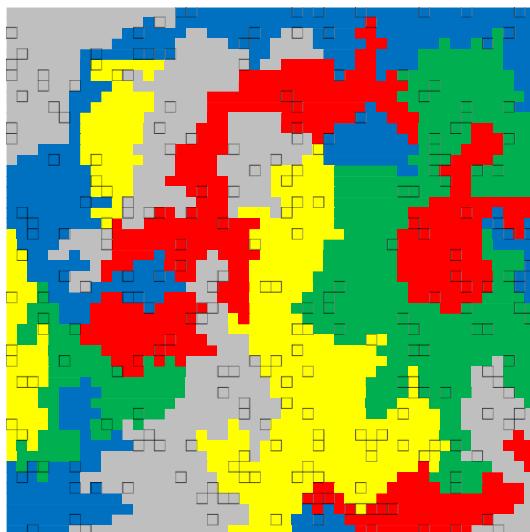
$$\text{Valor-P} = 3,95 \cdot 10^{-5}$$

CLASSIF.	REFERÊNCIA					n_u
	Red	Yellow	Green	Blue	Grey	
Red	360	0	110	0	0	470
Yellow	0	410	0	70	0	480
Green	0	90	390	0	50	530
Blue	80	0	0	370	0	450
Grey	60	0	0	60	450	570
n_u	500	500	500	500	500	2500

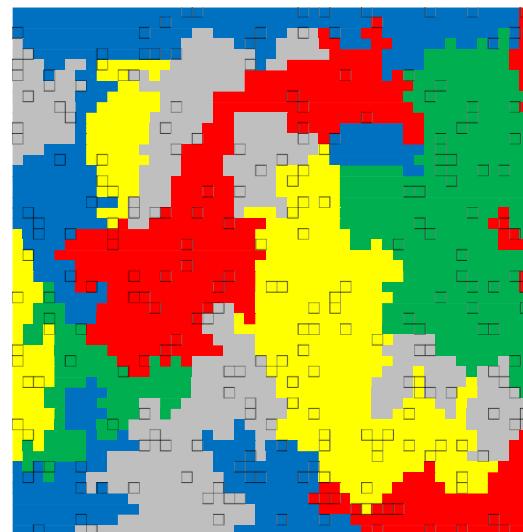
Conclusão: rejeita-se H_0 com 5% de significância, ou seja, o kappa dessa classificação é maior que 0,7.

(esta conclusão pode estar equivocada pois desconsiderou a dependência espacial)

Índice Kappa - Exemplo 3



classificação



referência

Selecionando-se
250 pontos...

$$H_0: \kappa = 0,7$$

$$H_1: \kappa > 0,7$$

$$\hat{\kappa} = 0,7336$$

$$\text{Var}(\hat{\kappa}) = 0,00103$$

Se H_0 for verdadeira, então:

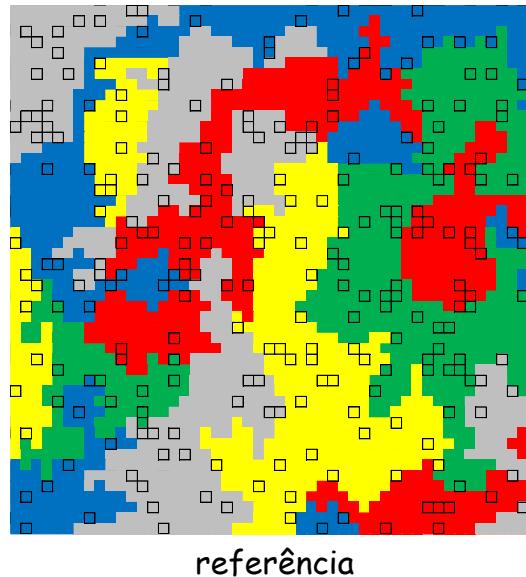
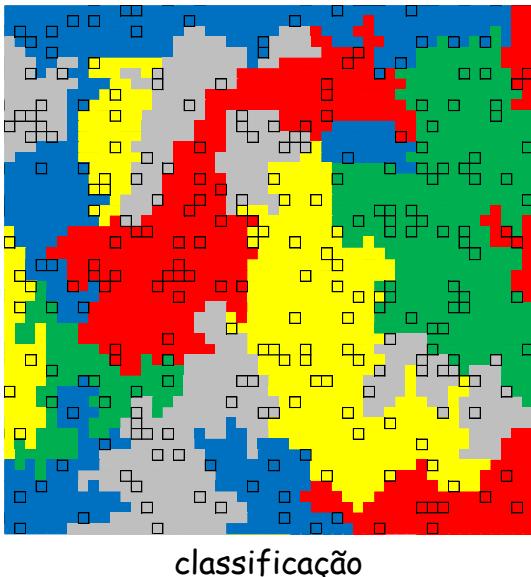
$$z = \frac{\hat{\kappa} - 0,7}{\sqrt{\text{Var}(\hat{\kappa})}} = 1,0447$$

$$\text{Valor-P} = 0,1481$$

CLASSIF.	REFERÊNCIA					n_i
	Red	Yellow	Green	Blue	Gray	
Red	26	0	8	0	0	34
Yellow	0	50	0	8	0	58
Green	0	8	35	0	6	49
Blue	14	0	0	33	0	47
Gray	6	0	0	3	53	62
n_i	46	58	43	44	59	250

Conclusão: aceita-se H_0 , ou seja, o valor de kappa observado não é significativamente (a 5%) maior que 0,7

Índice Kappa - Exemplo 4



Por que acreditar nesta única amostra de 250 pontos?

A cada nova amostragem de 250 pontos, uma nova matriz de confusão seria obtida e portanto um novo valor de kappa seria estimado...

$$H_0: \kappa = 0,7$$

$$H_1: \kappa > 0,7$$

Nova abordagem:

A escolha dos 250 pontos amostrais é repetido muitas vezes (pelo menos 1000 vezes).

Para cada amostragem, determina-se a matriz de confusão e calcula-se o kappa

Calcula-se a frequência (probabilidade) com que H_0 é aceita

Se esta frequência for baixa (menor que 5%, por exemplo), conclui-se que é relativamente "raro" aceitar H_0 e portanto rejeita-se H_0 .

Suponha que, após a simulação de 10000 valores de kappa, 10,78% desses valores são menores ou iguais a 0,7...

Conclusão: aceita-se H_0 , ou seja, não há fortes evidências que levem a conclusão que o kappa seja de fato maior que 0,7

Outros Índices

Kappa condicional

Avalia a concordância para uma determinada classe

Kappa ponderado

Cada célula ij da matriz de confusão pode receber um peso ($0 \leq w_{ij} \leq 1$) permitindo que certos erros sejam mais importantes que outros

Tau

O índice Kappa pressupõe que ambas as probabilidades marginais (classificação e referência) sejam conhecidas antes mesmo da classificação

Para o índice Tau, utiliza-se as probabilidades *a priori* de cada classe (p_k) ao invés de estimá-las pelas proporções marginais obtidas após a classificação

Assim, este índice pode ser obtido por:

$$\hat{\tau} = \frac{\theta_1 - \theta'_2}{1 - \theta'_2} \quad \theta_1 = \frac{\sum_{k=1}^c x_{kk}}{n} \quad \theta'_2 = \frac{\sum_{k=1}^c p_k x_{+k}}{n}$$

Na ausência de informação, utiliza-se o mesmo valor para todos p_k (classes equiprováveis, ou seja, $p_k = 1/c$)

Críticas ao Kappa*

- É sempre melhor (mais útil) focar na discordância e tentar explicar os erros do que focar na concordância e se preocupar como a aleatoriedade explica parte do acerto observado - o que é a essência do Kappa;
- Não se conhece nenhum artigo onde os autores mudaram a conclusão quando compararam a exatidão total e o Kappa. Em geral, os autores apenas apresentam simultaneamente a exatidão total e o Kappa junto ao mapa avaliado (**desaconselhável!**);
- A avaliação da significância do Kappa é feita pressupondo-se que parte da exatidão total observada é casual (Kappa = 0?). **Isso é quase sempre inútil;**
- Pontius e Millones (2011) e Pontius e Santacruz (2014) sugerem particionar os erros em diferentes componentes:
 - *Quantity* (quantidade, grandeza)
 - *Allocation* (alocação, atribuição):
 - *Exchange* (troca, permuta)
 - *Shift* (mudança, deslocamento)

} Estes índices são mais úteis quando empregados na análise de trajetórias (mapas obtidos em 2 datas)

*Pontius, R. G.; Millones, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment.
International Journal of Remote Sensing, 32(15): 4407-4429, 2011

Pontius, R. G.; Santacruz, A. Quantity, exchange, and shift components of difference in a square contingency table. International Journal of Remote Sensing, 35(21): 7543-7554, 2014

Índice Quantity

		Referência			
		A	B	C	D
Classificação	A	13	8	0	0
	B	8	10	0	3
	C	0	5	27	4
	D	0	0	0	32
	Total	21	23	27	39
		Total			
		110			

Observe que para a classe A, os erros de omissão são compensados pelos erros de inclusão.

Neste caso, a área estimada pela classificação é a mesma da referência, ou seja, a classificação poderia ser utilizada para estimar a **quantidade** da classe A presente na referência.

Índice Quantity

Classificação	Referência				Total
	A	B	C	D	
A	13	8	0	0	21
B	8	10	0	3	21
C	0	5	27	4	36
D	0	0	0	32	32
Total	21	23	27	39	110

Já para classe B, há 2 amostras a mais na referência do que na classificação, o que representa um erro de 1,8% em relação ao total.

Classe	Quantity
A	0,0%
B	1,8%
C	8,2%
D	6,4%
Total	8,2%

Considerando todas as classes, $Quantity = 8,2\%$

$$Q_T = \frac{Q_A + Q_B + Q_C + Q_D}{2}$$

Índice Allocation

		Referência				
		A	B	C	D	Total
Classificação	A	13	8	0	0	21
	B	8	10	0	3	21
	C	0	5	27	4	36
	D	0	0	0	32	32
	Total	21	23	27	39	110

Em toda a matriz observa-se um total de 28 amostras erradas, o que representa 25,5%, ou seja, 100% - exatidão.

Como o erro devido a quantidade foi de 8,2%, então o restante, 17,3%, é devido a problemas de atribuição.

Mas qual tipo de erro de atribuição? Permuta ou mudança?

Índice Allocation: Exchange or Shift?

Classificação	Referência				Total
	A	B	C	D	
A	13	8	0	0	21
B	8	10	0	3	21
C	0	5	27	4	36
D	0	0	0	32	32
Total	21	23	27	39	110

Observe que para a classe A, todos os erros foram devidos à permuta com a classe B, ou seja, amostras classificadas como B foram compensadas pelas amostras da classe B erroneamente classificadas como A. Dessa forma, para a classe A, todo o erro de atribuição foi do tipo *exchange* (permuta) totalizando 16 amostras ($8 + 8$) ou 14,5% do total.

Índice Allocation: Exchange or Shift?

		Referência				
		A	B	C	D	Total
Classificação	A	13	8	0	0	21
	B	8	10	0	3	21
	C	0	5	27	4	36
	D	0	0	0	32	32
	Total	21	23	27	39	110

Já a classe B, 8 amostras erradas foram permutadas com a classe A (*Exchange* = 14,5%), mas 5 foram erroneamente classificadas como classe C e 3 foram erroneamente alocadas para a classe B sendo verdadeiramente da classe D.

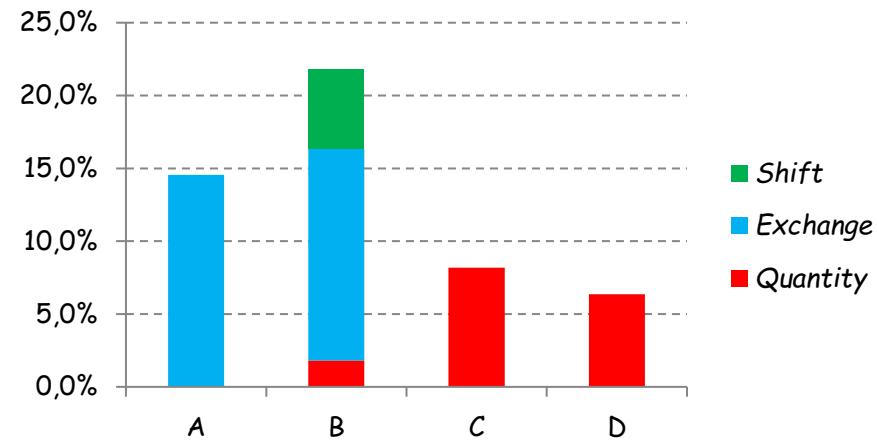
Nesse caso, dos 24 erros observados, houve 2 erros de quantidade (*Quantity* = 1,8%), 16 erros de permuta (*Exchange* = 14,5%), restando 6 erros por mudança (*Shift* = 5,5%).

Índices propostos por Pontius

Classificação	Referência				Total
	A	B	C	D	
A	13	8	0	0	21
B	8	10	0	3	21
C	0	5	27	4	36
D	0	0	0	32	32
Total	21	23	27	39	110

Classe	Quantity Allocation	
A	0,0%	14,5%
B	1,8%	20,0%
C	8,2%	0,0%
D	6,4%	0,0%
Total	8,2%	17,3%

Exchange	Shift
14,5%	0,0%
14,5%	5,5%
0,0%	0,0%
0,0%	0,0%
14,5%	2,7%



Avaliação da Exatidão - Comentários Gerais

- A Matriz de Confusão constitui a parte mais importante na avaliação desde que represente corretamente os acertos e erros entre o mapa e a referência
- O Kappa e outros índices calculados para um mapa específico tem pouco (ou nenhuma) utilidade prática. Esses índices são úteis apenas quando comparados a outros resultados. **Nunca** use tabelas que qualificam valores de Kappa.
~~Ex.: Segundo L & K (1977), o kappa de 0,7934 é considerado bom...~~
- Ainda não há uma avaliação da significância dos índices propostos por Pontius, o que dificulta a comparação de resultados entre diferentes métodos de classificação
- Em geral, as avaliações da exatidão não consideram a componente espacial

Avaliação da Exatidão - Comentários Gerais

- Os erros encontrados deveriam apontar para melhoria no método de classificação
- Comumente, a coleta das amostras de referência é feita com base no mapa resultante da classificação. Se as amostras não forem distribuídas proporcionalmente a cada classe, índices relacionados ao produtor (referência) podem ser enviesados. Pontius e Millones (2011) indicam sempre utilizar a matriz não enviesada (ajustada para as **proporções reais** de cada classe na classificação)
- As técnicas de reamostragem poderiam ser melhor aproveitadas para avaliações das incertezas, principalmente quando se dispõe de uma referência “completa” (por que confiar nos resultados de uma única amostragem?)

Avaliação de Classificação Binária

Em alguns casos, as classificações admitem apenas 2 classes: presente/ausente, ocorre/não ocorre, sim/não, positivo/negativo, etc

Nesse caso, a matriz de confusão teria a forma:

		Referência	
		Positivo	Negativo
Classificação	Positivo	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Negativo	Falso Negativo (FN)	Verdadeiro Negativo (VN)

$$\text{exatidão} = \frac{VP + VN}{VP + FP + FN + VN}$$

$$\text{sensibilidade} = \frac{VP}{VP + FN}$$

$$\text{precisão} = \frac{VP}{VP + FP}$$

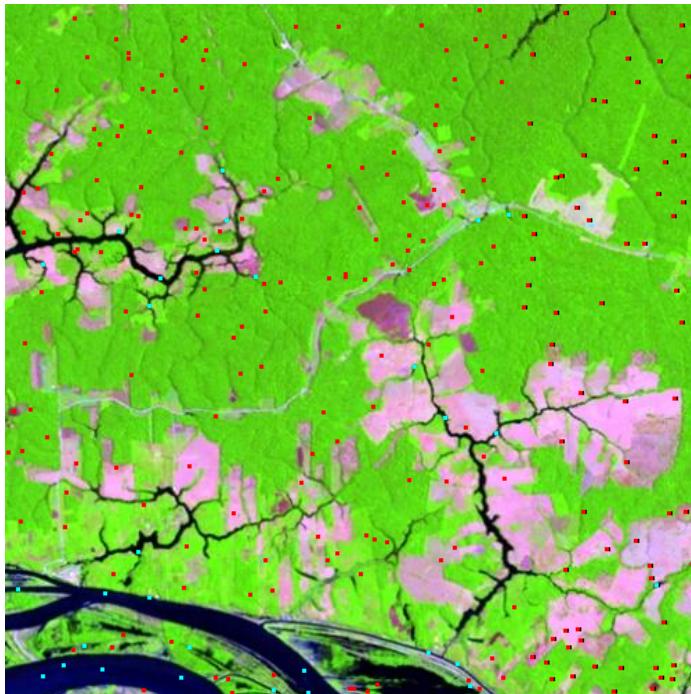
$$F1\ score = 2 \frac{\text{precisão} * \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}}$$

O *F1 score* é representa uma média balanceada entre a precisão e a sensibilidade

OBS: sensibilidade = revocação = *recall* = taxa de verdadeiros positivos - equivale à exatidão do produtor da classe positivo
precisão - equivale à exatidão do usuário da classe positivo

Avaliação de Classificação Binária

Exemplo: Detecção de corpos d'água

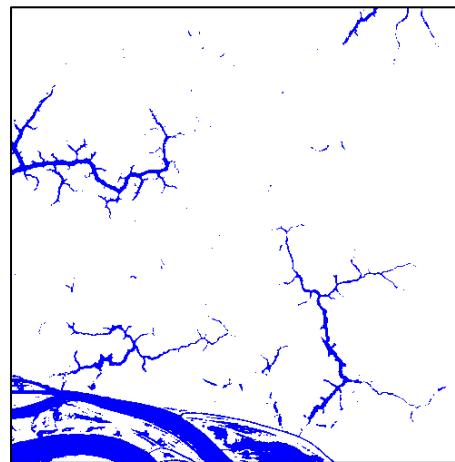


Landsat TM 5R4G3B

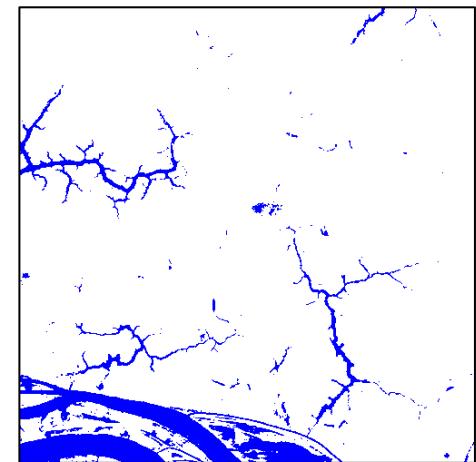
Amostras de treinamento:

- 30 de água
- 230 de não água

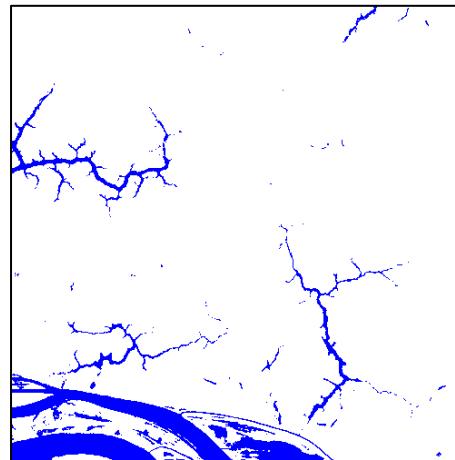
Qual teve melhor desempenho?



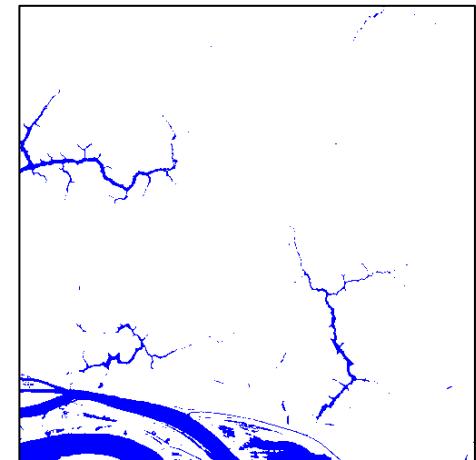
Mínima Distância



Mahalanobis



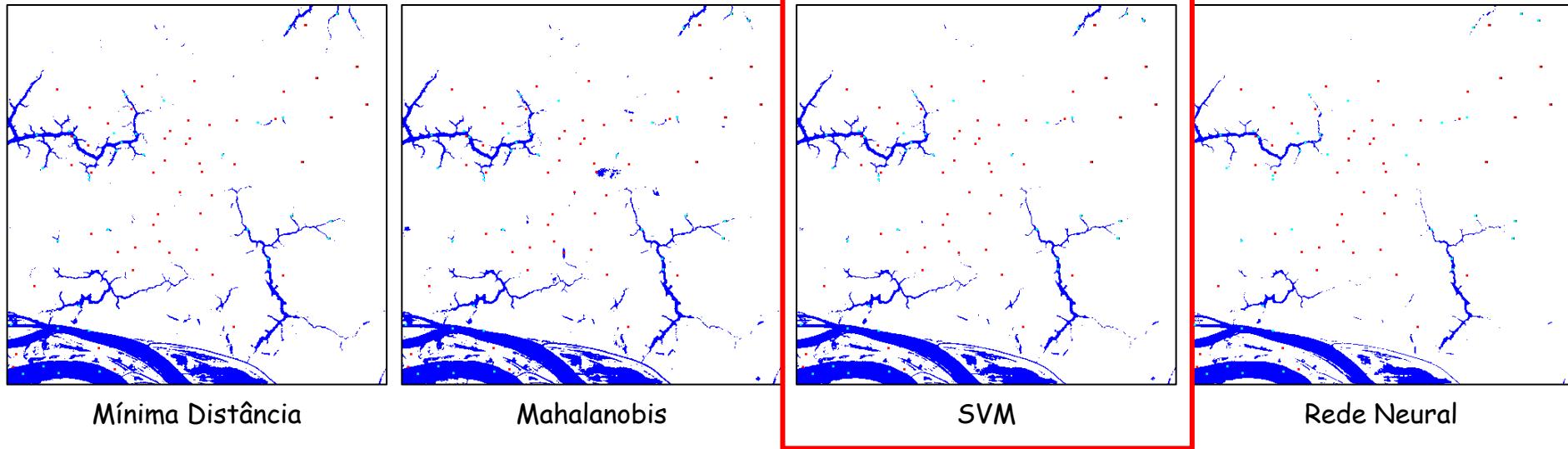
SVM



Rede Neural

Avaliação de Classificação Binária

Exemplo: Detecção de corpos d'água

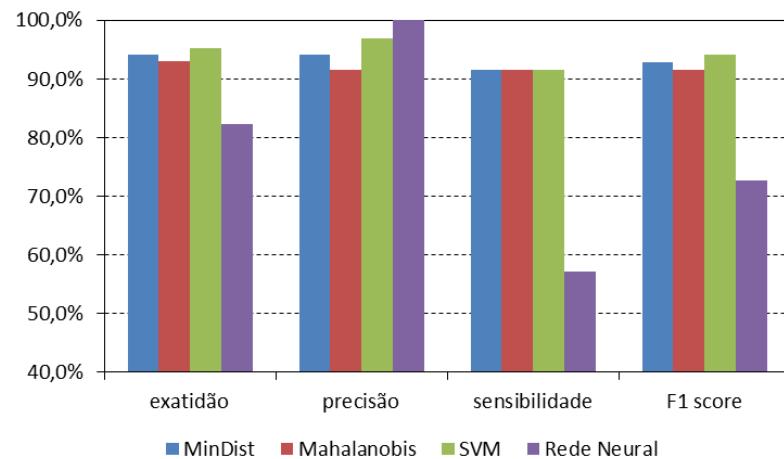


Amostras de teste:

- 35 de água
- 50 de não água

Class	MDist	Ref	
		Água	NÁgua
Água	32	2	
NÁgua	3	48	

Class	Mahal	Ref	
		Água	NÁgua
Água	32	3	
NÁgua	3	47	



Class	SVM	Ref	
		Água	NÁgua
Água	32	1	
NÁgua	3	49	

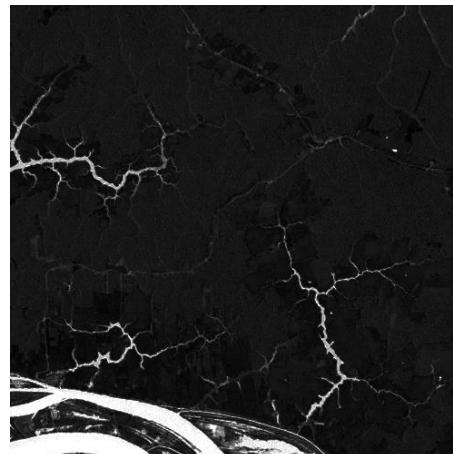
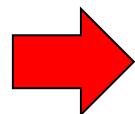
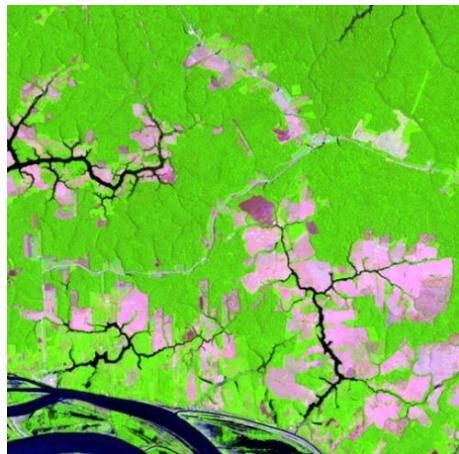
Class	RN	Ref	
		Água	NÁgua
Água	20	0	
NÁgua	15	50	

Geração de Mapas Binários por Fatiamento

Caso seja necessário a produção de um mapa a partir de uma métrica ou índice que representa uma probabilidade ou risco de ocorrência, é imprescindível a adoção de um limiar que irá definir a faixa de valores associados a classe desejada. Isso acontece em classificadores fuzzy, regressões logísticas e modelos de riscos em geral.

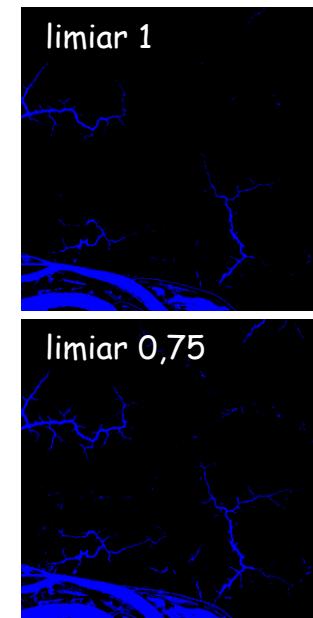
Essa mesma ideia pode ser aplicada quando se utiliza um índice espectral que maximiza a detecção de um determinação alvo (índices de vegetação, de água, de área construída, etc).

Mas como definir o melhor limiar nesse caso?



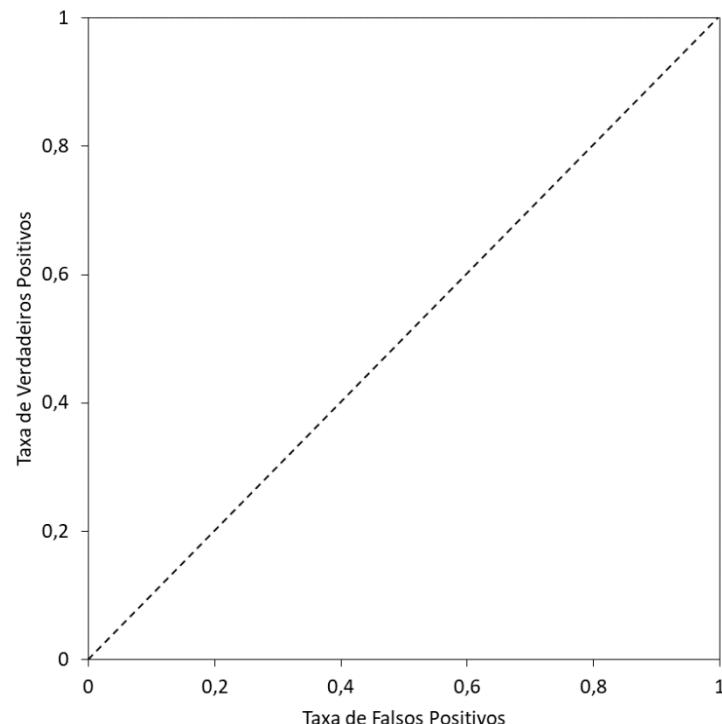
Landsat TM 5R4G3B

B3/B5



Curva ROC

O gráfico ROC (*Receiver Operating Characteristics*) pode ser usado para avaliar e comparar o desempenho de classificadores, ou definir o melhor limiar a ser utilizado. Permite ponderar a escolha entre benefícios (verdadeiros positivos) e custos (falsos positivos).

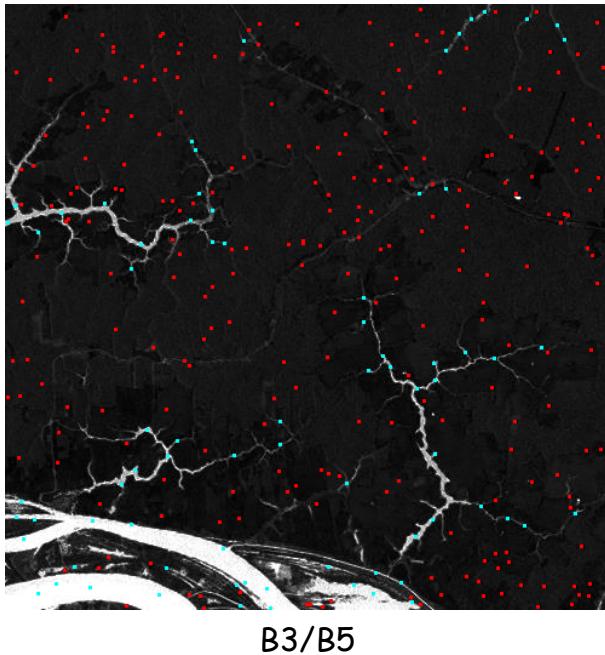


$$\text{taxa } VP = \frac{VP}{VP + FN} \quad (\text{taxa } VP = \text{sensibilidade} = \text{recall})$$

$$\text{taxa } FP = \frac{FP}{FP + VN} \quad (1 - \text{taxa } FP = \text{especificidade})$$

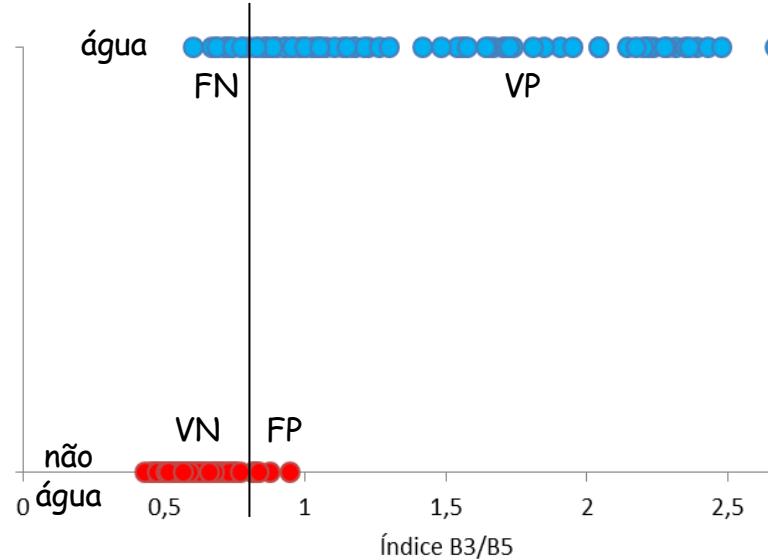
Quanto mais o resultado se aproximar do canto superior esquerdo (posição 0,1) melhor seu desempenho
A linha tracejada indica que o resultado representa uma estratégia aleatória, sem privilegiar os verdadeiros positivos em detrimento aos falsos positivos

Curva ROC



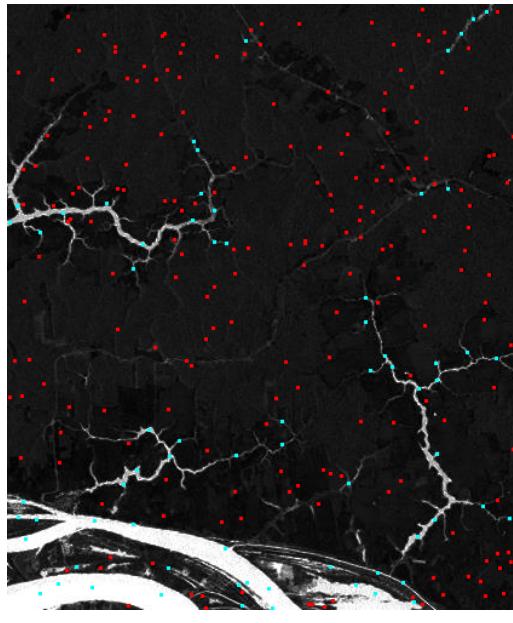
Amostras de treinamento:

- 75 de água
- 250 de não água

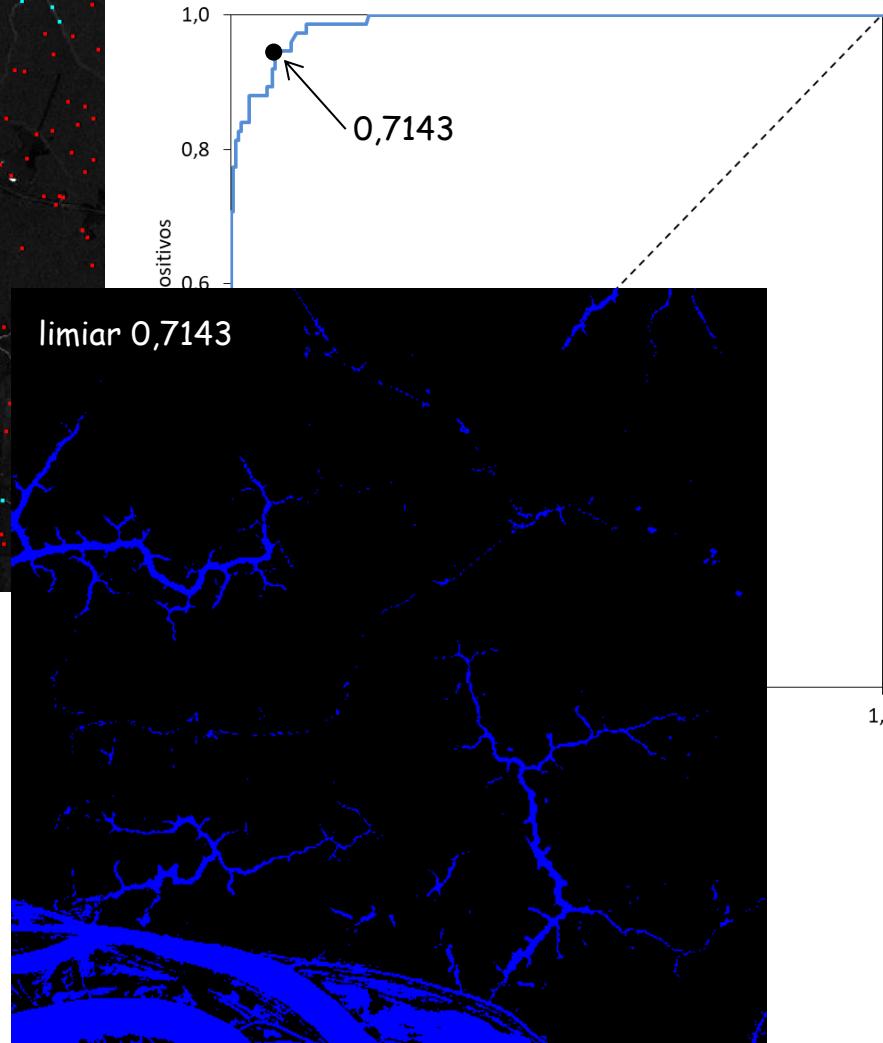


Índice	TVP	TFP
< 0,4316	1,000	1,000
0,4316	1,000	1,000
0,4476	1,000	0,996
0,4600	1,000	0,992
0,4615	1,000	0,988
0,4706	1,000	0,984
0,4715	1,000	0,980
0,4745	1,000	0,976
0,4762	1,000	0,972
0,4767	1,000	0,968
:	:	:
2,2778	0,120	0,000
2,2857	0,107	0,000
2,3158	0,093	0,000
2,3500	0,080	0,000
2,3636	0,067	0,000
2,3913	0,053	0,000
2,4286	0,040	0,000
2,4783	0,027	0,000
2,6667	0,013	0,000
>2,6667	0,000	0,000

Curva ROC

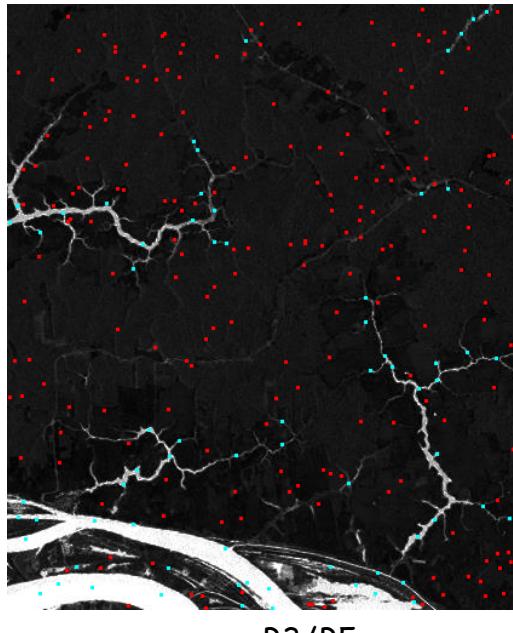


Amostras de treinamento:
● 75 de água
● 250 de não água

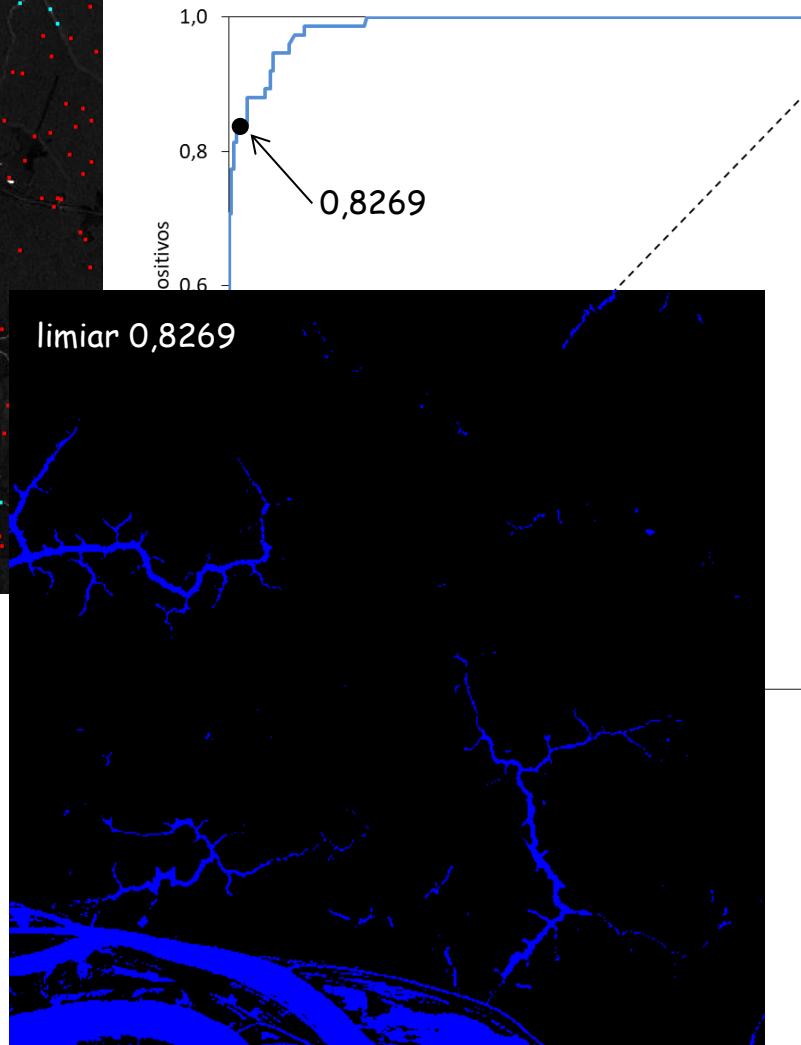


Índice	TVP	TFP
< 0,4316	1,000	1,000
0,4316	1,000	1,000
0,4476	1,000	0,996
0,4600	1,000	0,992
0,4615	1,000	0,988
0,4706	1,000	0,984
0,4715	1,000	0,980
0,4745	1,000	0,976
0,4762	1,000	0,972
0,4767	1,000	0,968
:	:	:
2,2778	0,120	0,000
2,2857	0,107	0,000
2,3158	0,093	0,000
2,3500	0,080	0,000
2,3636	0,067	0,000
2,3913	0,053	0,000
2,4286	0,040	0,000
2,4783	0,027	0,000
2,6667	0,013	0,000
>2,6667	0,000	0,000

Curva ROC



Amostras de treinamento:
● 75 de água
● 250 de não água



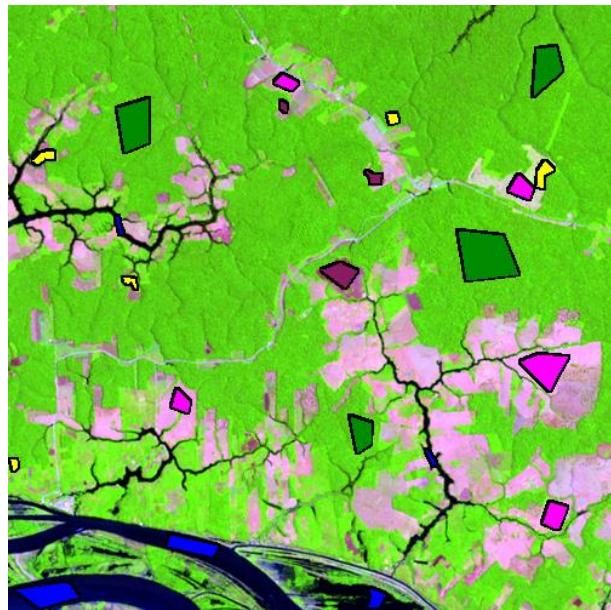
Índice	TVP	TFP
< 0,4316	1,000	1,000
0,4316	1,000	1,000
0,4476	1,000	0,996
0,4600	1,000	0,992
0,4615	1,000	0,988
0,4706	1,000	0,984
0,4715	1,000	0,980
0,4745	1,000	0,976
0,4762	1,000	0,972
0,4767	1,000	0,968
⋮	⋮	⋮
2,2778	0,120	0,000
2,2857	0,107	0,000
2,3158	0,093	0,000
2,3500	0,080	0,000
2,3636	0,067	0,000
2,3913	0,053	0,000
2,4286	0,040	0,000
2,4783	0,027	0,000
2,6667	0,013	0,000
>2,6667	0,000	0,000

Mapas de Incerteza

Classificadores Probabilísticos

Os classificadores probabilísticos utilizam do conhecimento prévio da distribuição dos valores dos atributos referentes às classes de interesse para definir qual é a classe mais provável que um determinado elemento (ainda desconhecido) deve pertencer

Em geral, estas distribuições são desconhecidas e devem ser estimadas através de amostras (de treinamento)



Classes

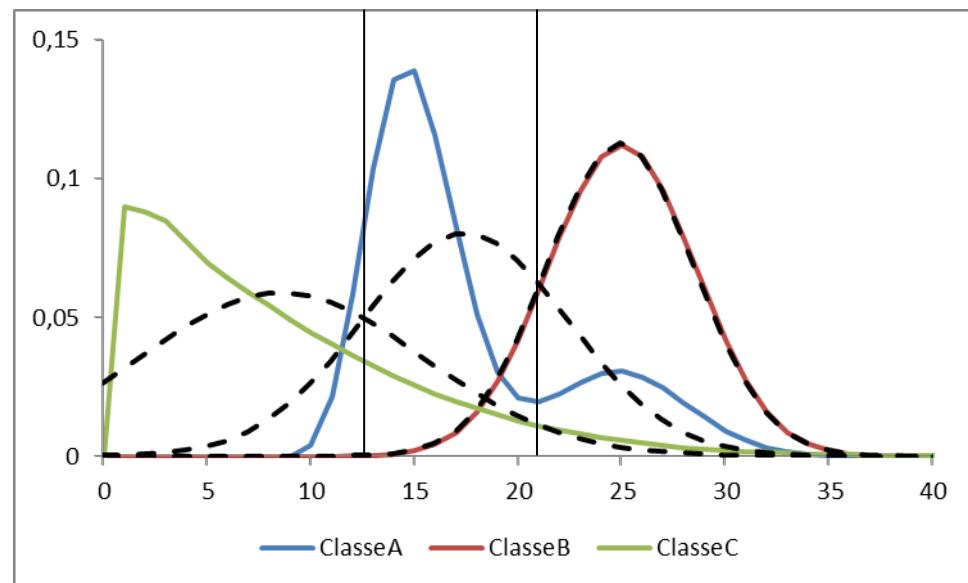
- Floresta
- Regeneração
- Desmatamento
- Queimada
- Água

A definição das classes de interesse é etapa fundamental neste processo

Nesse caso, é importante conhecer o tipo de distribuição que o classificador é capaz de modelar!

Classificadores Probabilísticos

Um erro comum é confundir o que se quer classificar com a capacidade do classificador em representar a distribuição dos valores reais presentes na imagem



Supondo um classificador gaussiano...

Nos casos em que a distribuição parece ser bimodal (ClasseA), provavelmente há subclasses que devem ser separadas, o que resultaria num melhor o desempenho do classificador

Dessa forma, deve-se definir muito mais classes do que o que se deseja e posteriormente estas classes podem ser agrupadas.

Exemplo: presença de nuvens!

Se a distribuição diferir muito da distribuição suposta válida, deve-se buscar outros classificadores...

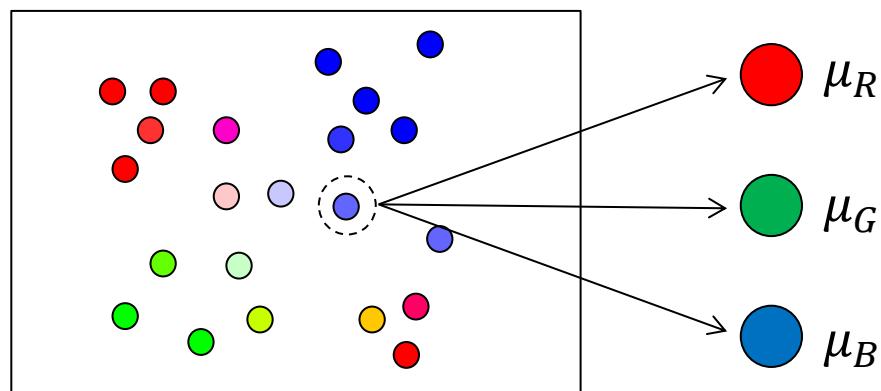
Classificadores Fuzzy

Define-se um classificador fuzzy como aquele que utiliza conjuntos fuzzy ou lógica fuzzy durante sua criação ou operação

Na classificação fuzzy, para cada elemento define-se o grau de pertinência para cada classe K (rotulação soft). Um classificador probabilístico também pode produzir uma rotulação soft, mas seu significado é diferente

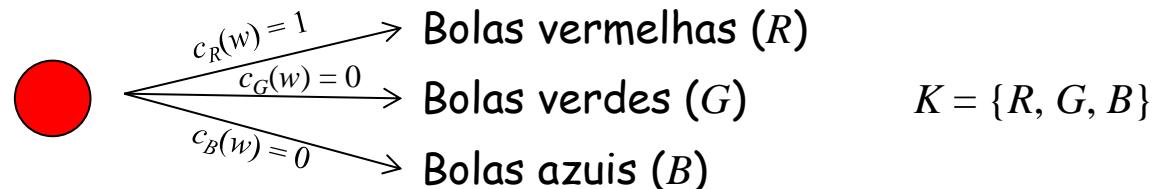
Basicamente pode-se dividir os classificadores fuzzy em 2 tipos:

- baseados em regras fuzzy
- os que consideram um protótipo fuzzy

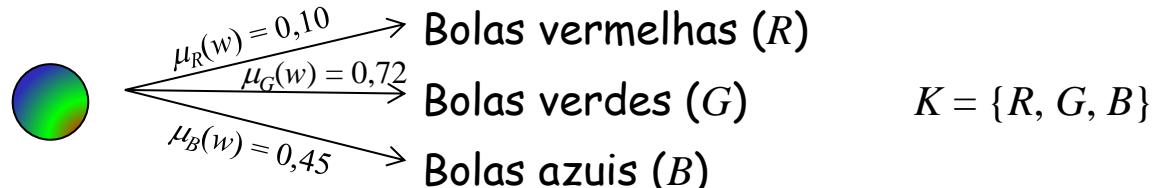


Teoria Fuzzy ("nebulosa")

Na teoria clássica, a **função característica** de um elemento w a um conjunto K , $c_K(w)$ só admite valores 0 ou 1



Na teoria fuzzy, é feita uma generalização da função característica, dando origem a chamada **função de pertinência** de um elemento w a um conjunto K , $\mu_K(w)$, que permite associar a cada elemento graus de pertinência a K . O contradomínio desta função pode ser qualquer um mas, normalmente, por simplicidade e convenção, é considerado o intervalo $[0,1]$

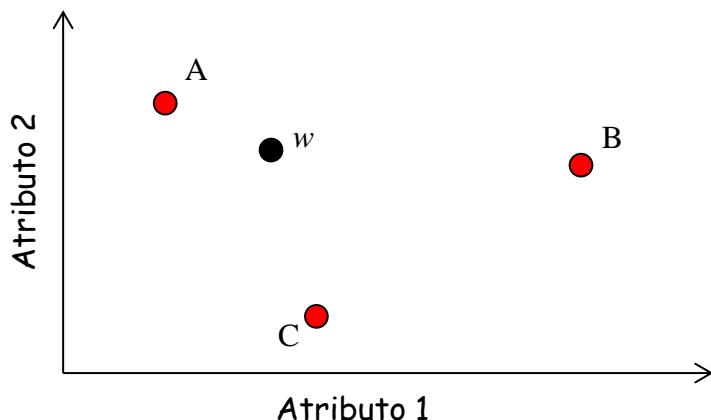


OBS: A função de pertinência, ao contrário da função de probabilidade, não precisa somar 1, ou seja, não representam proporções.

Teoria Fuzzy ("nebulosa")

Segundo Dubois e Prade (1997), pode-se definir três interpretações para os graus de pertinência: graus de semelhança, graus de preferência e graus de incerteza

Para classificação, a interpretação mais usual é o **grau de semelhança** utilizando-se medidas de similaridade ou medidas de proximidade (distâncias) em relação a um elemento típico representativo de uma determinada classe. Esta interpretação é utilizada, por exemplo, em análise de agrupamento (*cluster*)



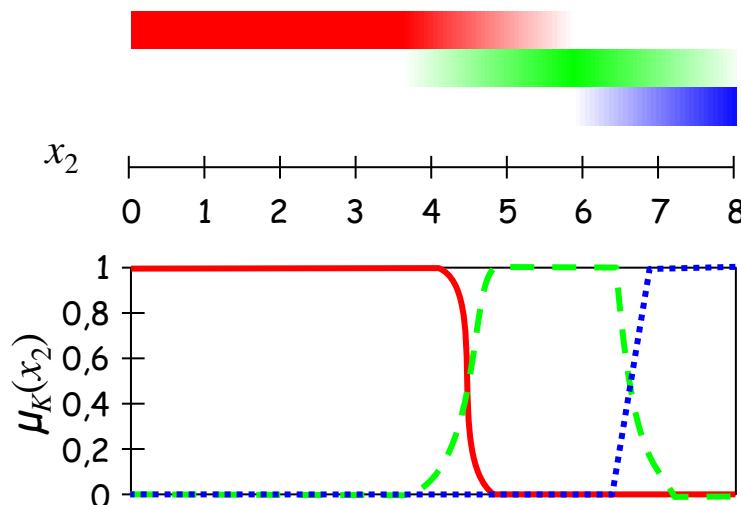
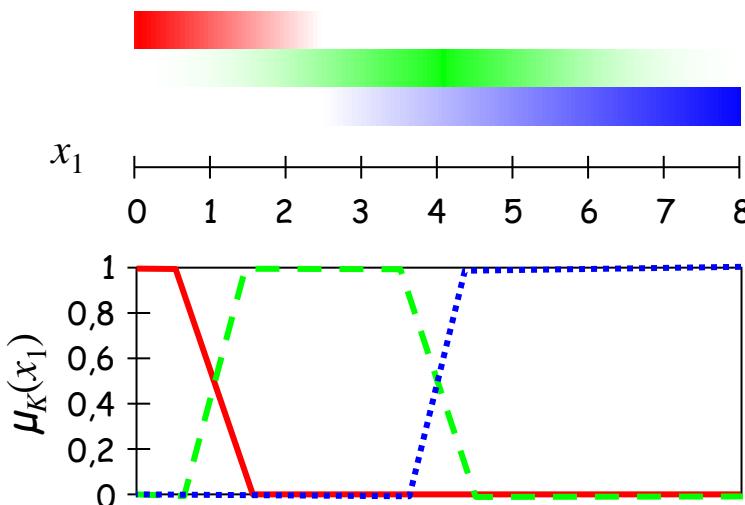
Se for analisar a distância entre w e os pontos A, B e C, pode-se concluir que:

w se parece mais com A, um pouco com C e muito pouco com B

Classificação Fuzzy - Regras Fuzzy

Esse classificador baseia-se em regras do tipo SE-ENTÃO.

Há diversos subtipos desse classificador. Aqui abordaremos o mais simples deles, cujo resultado da aplicação da regra é um rótulo apenas, usando um exemplo com 2 variáveis numéricas mas representadas nas regras na forma categórica (cor):



As transições podem seguir diferentes funções: linear, sigmóide, quadrática, exponencial, etc

Classificação Fuzzy - Regras Fuzzy

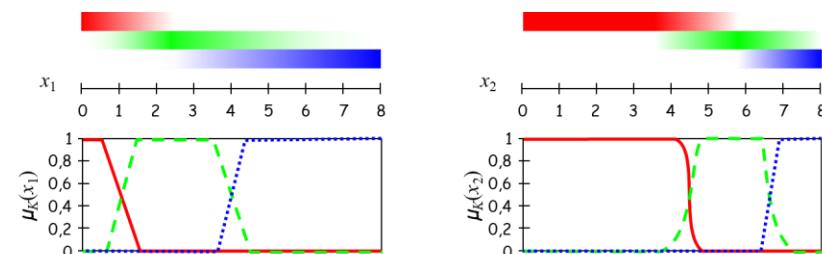
Em seguida, definem-se as regras:

SE x_1 é verde E x_2 é vermelho ENTÃO classe é 1

SE x_1 é verde E x_2 é azul ENTÃO classe é 2

SE x_1 é azul E x_2 é vermelho ENTÃO classe é 3

SE x_1 é vermelho E x_2 é azul ENTÃO classe é 4



E a força (ou peso) de cada regra

Por exemplo, para a 1^a regra, considerando que o elemento w possui os valores

$\mathbf{x} = (x_1, x_2)$, tem-se:

$$\tau_1(\mathbf{x}) = \mu_G(x_1) \text{ E } \mu_R(x_2)$$

(operador de mínimo)

Calculado a força de todas as regras, define-se o rótulo *soft* de cada elemento como:

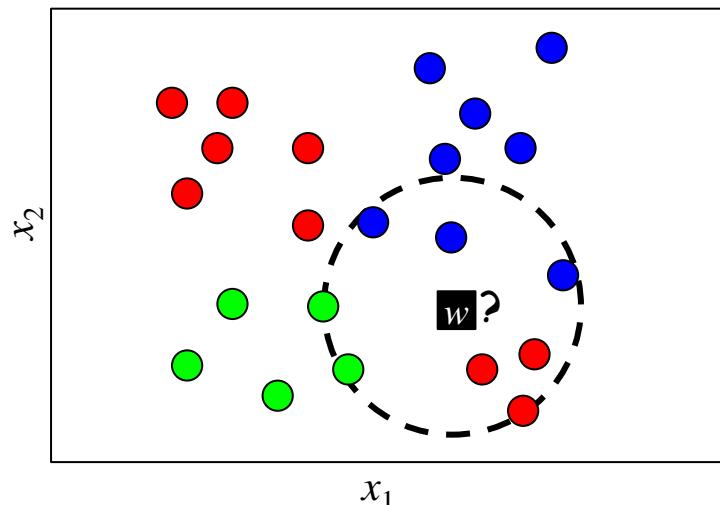
$$g(\mathbf{x}) = \max(\tau_i(\mathbf{x}))$$

Para uma classificação *crisp*, pode-se considerar a classe com maior valor de $g(\mathbf{x})$

Classificação Fuzzy - Protótipo Fuzzy

Muitos classificadores fuzzy são inspirados na ideia de se “fuzzificar” classificadores tradicionais

Um típico representante deste grupo é o classificador K-vizinhos mais próximos (K-nn). No K-nn clássico, o elemento w é rotulado na classe melhor representada entre seus k vizinhos mais próximos (classe majoritária)

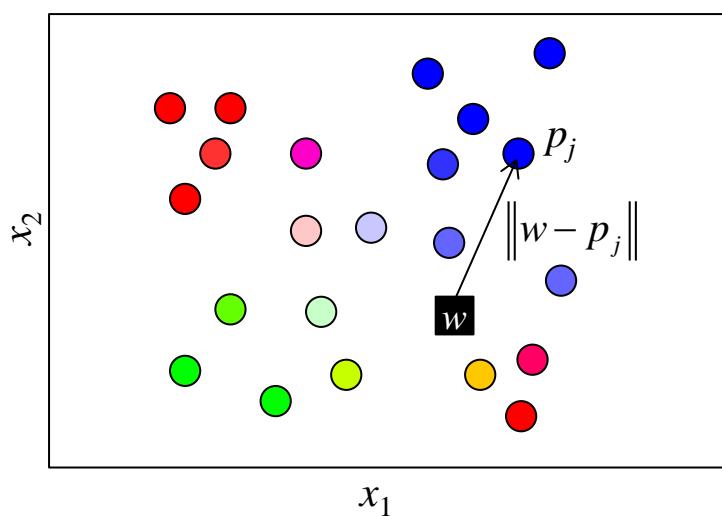


Nesse caso, pode-se obter uma estimativa grosseira da probabilidade *a posteriori* através do cálculo da proporção de vizinhos que possuem a classe designada para w

Classificação Fuzzy - Protótipo Fuzzy

Numa versão fuzzy, este classificador pode utilizar a informação de distância entre os vizinhos durante o processo de classificação.

O rótulo *soft* de cada elemento w pode ser obtido através da combinação das pertinências de cada um dos protótipos p_j ($j = 1, \dots, N$) considerando-se as distâncias relativas a w



$$\mu_K(w) = \frac{\sum_{j=1}^N \mu_K(p_j) \left(\frac{1}{\|w - p_j\|^{2/(m-1)}} \right)}{\sum_{j=1}^N \left(\frac{1}{\|w - p_j\|^{2/(m-1)}} \right)}$$
$$m > 1$$

Os protótipos utilizados como referência não precisam ser selecionados a partir dos dados existentes. Um conjunto de protótipos podem ser construídos a partir de conhecimento prévio.

Mapas de Incerteza

Usualmente a avaliação de uma classificação é feita através de índices globais (exatidão global, kappa, tau, etc) que refletem uma condição geral de acertos e erros mas podem esconder defeitos e qualidades locais

As incertezas oriundas da classificação são quase sempre ignoradas durante o processo de avaliação de um mapa temático

Mas como avaliar as incertezas espacialmente distribuídas?

- Definição explícita das incertezas do classificador: através da comparação das regras de decisão que levaram a escolha de uma dentre as várias classes possíveis (rotulação soft)
- Simulação Monte Carlo (quando não se tem acesso às regras de decisão): através da replicação do processo de classificação, a partir de diferentes amostras, gerando muitos mapas possíveis e contabilizando a frequência com que cada classe ocorreu nas diversas simulações

Esta abordagem é semelhante a feita pelo classificador Random Forest

Medidas de Incerteza

Considerando um elemento w que pode ser atribuído a classe K_i dentre as C classes possíveis, podemos utilizar as seguintes medidas para representar a incerteza desta atribuição:

Baseando-se na **função de probabilidade**

- Entropia de Shannon

$$H(w) = - \sum_{i=1}^C P(w \in K_i) \log_2 P(w \in K_i) \quad 0 \leq H(w) \leq \log_2 C \quad (\text{o log pode ter outras bases})$$

- Razão de Incerteza

$$RI(w) = 1 - \frac{\max[P(w \in K_i)] - \frac{1}{C}}{1 - \frac{1}{C}} \quad 0 \leq RI(w) \leq 1$$

- Complementar da Probabilidade Máxima

$$Cmax(w) = 1 - \max[P(w \in K_i)] \quad 0 < Cmax(w) \leq 1$$

- Razão entre 1º e 2º colocados (Índice de Confusão)

$$IC(w) = P(w \in K_2) / P(w \in K_1) \quad 0 < IC(w) \leq 1 \quad P(w \in K_1) \geq P(w \in K_2) \geq \dots \geq P(w \in K_C)$$

Medidas de Incerteza

Considerando um elemento w que pode ser atribuído a classe K_i dentre as C classes possíveis, podemos utilizar as seguintes medidas para representar a incerteza desta atribuição:

Baseando-se na **função de pertinência**

- **Incerteza U**

$$U(w) = [1 - \mu_1(w)] \log_2 C + \sum_{i=1}^{C-1} [\mu_i(w) - \mu_{i+1}(w)] \log_2 i \quad 0 \leq U(w) \leq \log_2 C$$
$$\mu_1(w) \geq \mu_2(w) \geq \dots \geq \mu_C(w)$$

- **Não-especificidade**

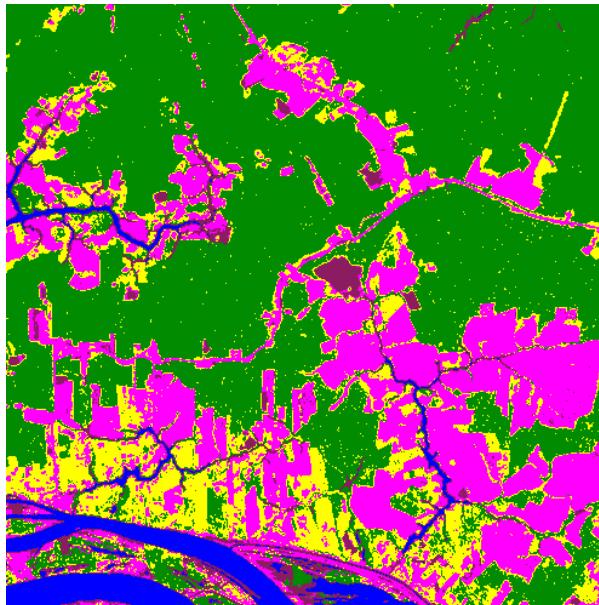
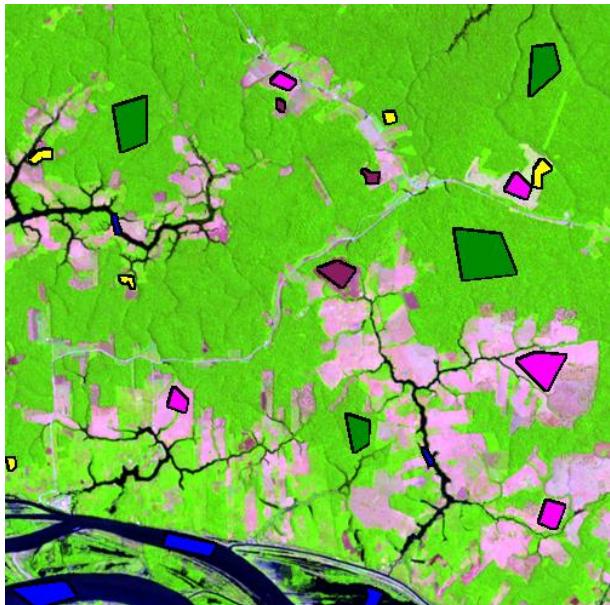
$$NSp(w) = 1 - \sum_{i=1}^C [\mu_i(w) - \mu_{i+1}(w)] \frac{1}{i} \quad 0 \leq NSp(w) \leq 1$$

- **Razão de Incerteza**

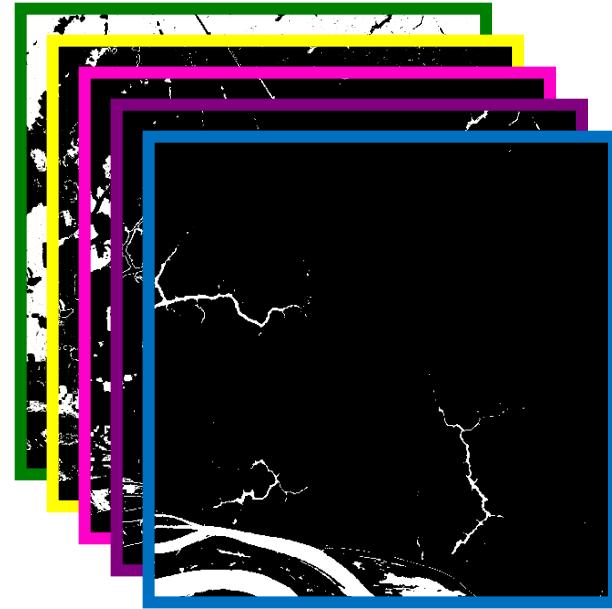
$$RI(w) = 1 - \frac{\max[\mu_i(w)] - \frac{\sum_{i=1}^N \mu_i(w)}{C}}{1 - \frac{1}{C}} \quad 0 \leq RI(w) \leq 1^*$$
$$* \max(RI) = 1 \text{ se } 0 \leq \mu_i(w) \leq 1 \text{ e } \sum \mu_i(w) = 1$$

Mapas de Incerteza - Exemplo 1

LANDSAT 7 ETM O/P 230/62 RGB543 08 set 2002



Classificação MaxVer



Grades de Probabilidades

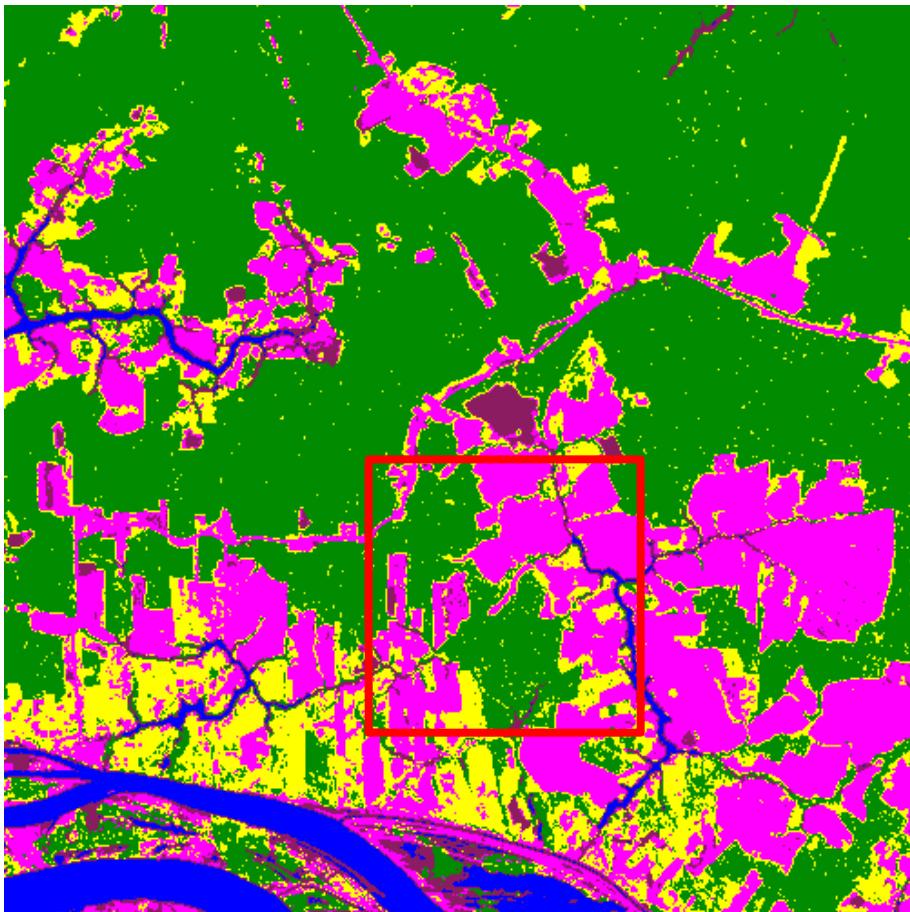
Classes

- Floresta (3262 pixels)
- Regeneração (379 pixels)
- Desmatamento (1371 pixels)
- Queimada (392 pixels)
- Água (1186 pixels)

Para diminuir os efeitos negativos da autocorrelação espacial, escolheu-se aleatoriamente apenas 10% dos pontos por classe
Estimam-se as médias, as variâncias/covariâncias das bandas para cada classe (ajuste de uma distribuição normal n-dimensional)
Calcula-se a probabilidade de cada pixel pertencer a cada classe
Mapa Final: classe com maior probabilidade

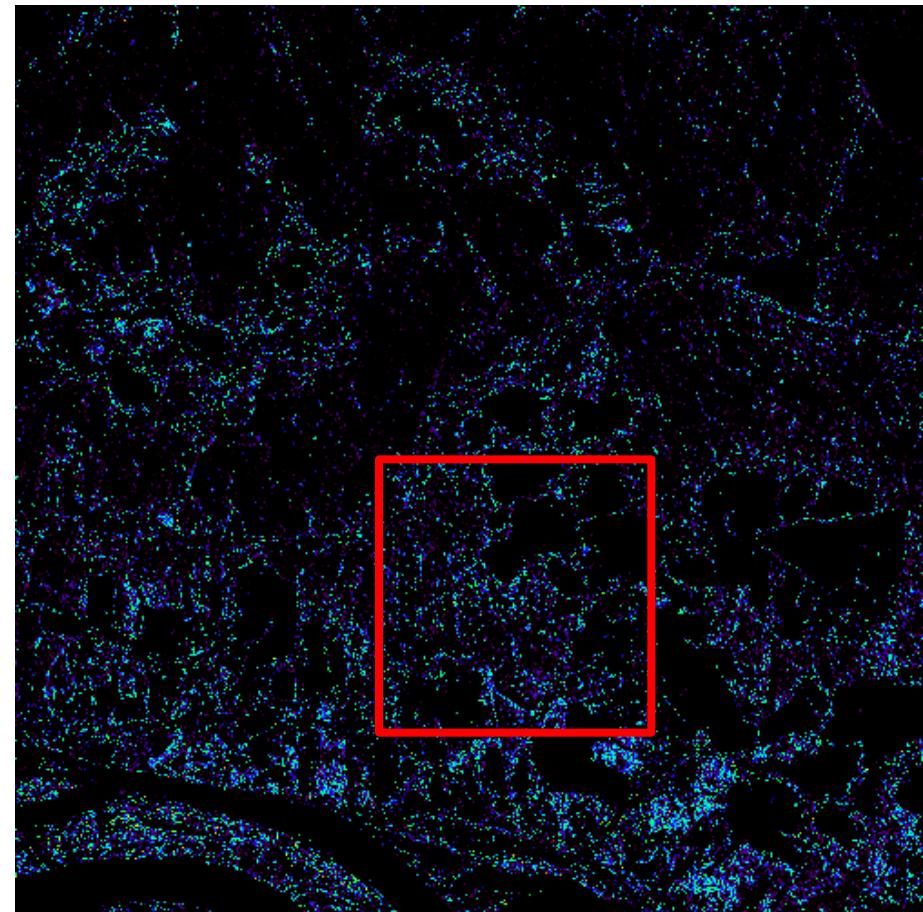
Classificador MaxVer Gaussiano

Mapas de Incerteza - Exemplo 1



Classificação MaxVer

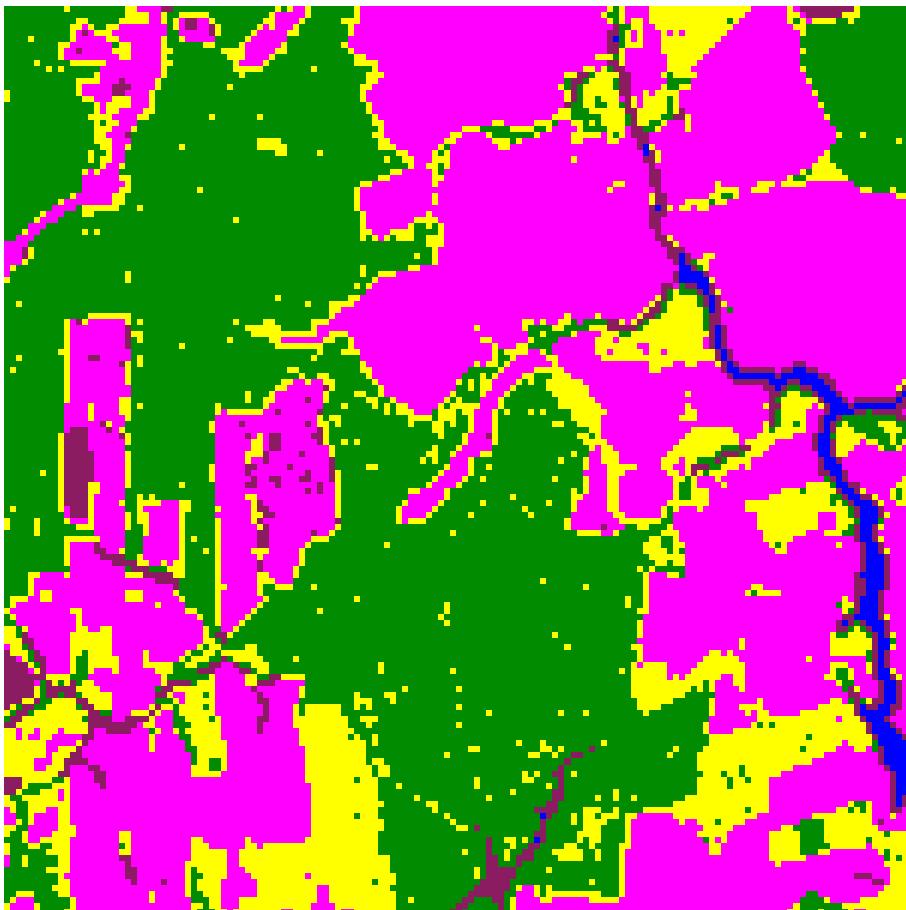
- | | |
|--------------|----------|
| Floresta | Queimada |
| Regeneração | Água |
| Desmatamento | |



Entropia de Shannon

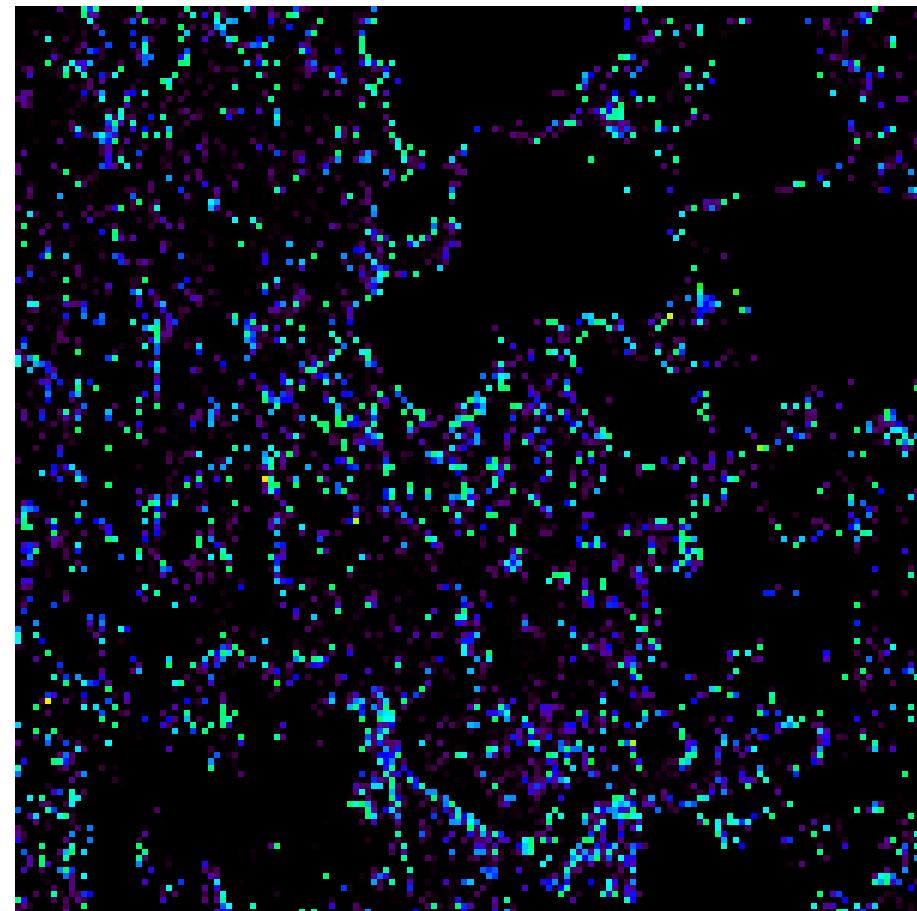


Mapas de Incerteza - Exemplo 1



Classificação MaxVer

- | | |
|--------------|----------|
| Floresta | Queimada |
| Regeneração | Água |
| Desmatamento | |



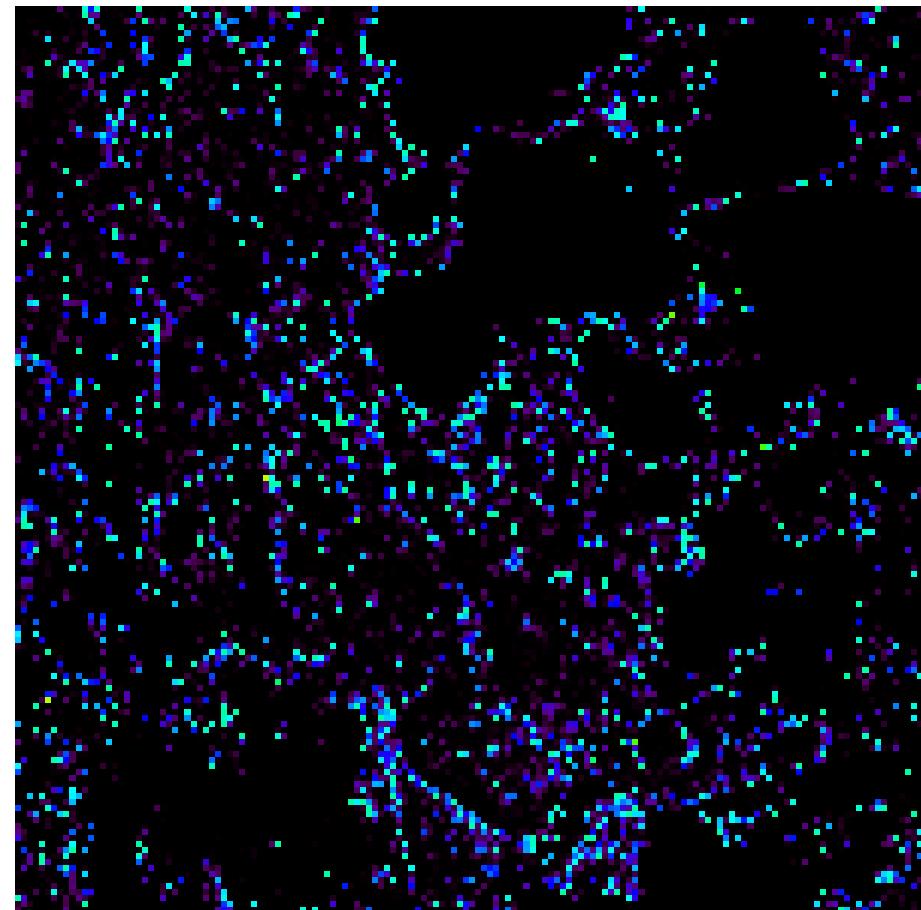
Entropia de Shannon



Mapas de Incerteza - Exemplo 1



Imagen TM RGB543

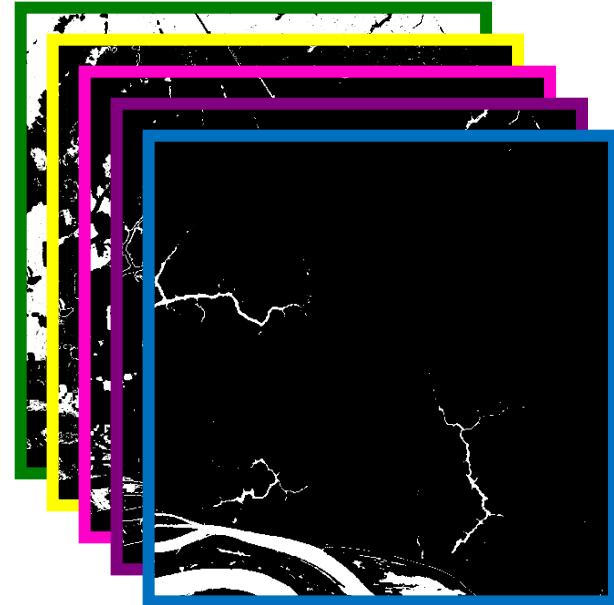
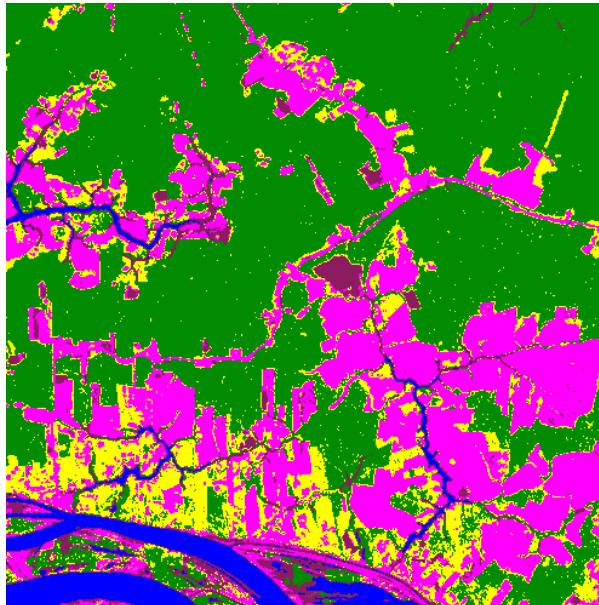
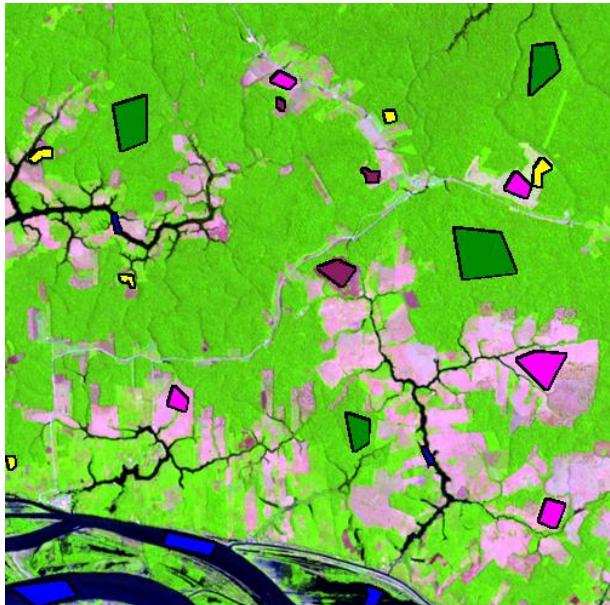


Entropia de Shannon



Mapas de Incerteza - Exemplo 2

LANDSAT 7 ETM O/P 230/62 RGB543 08 set 2002



Classes

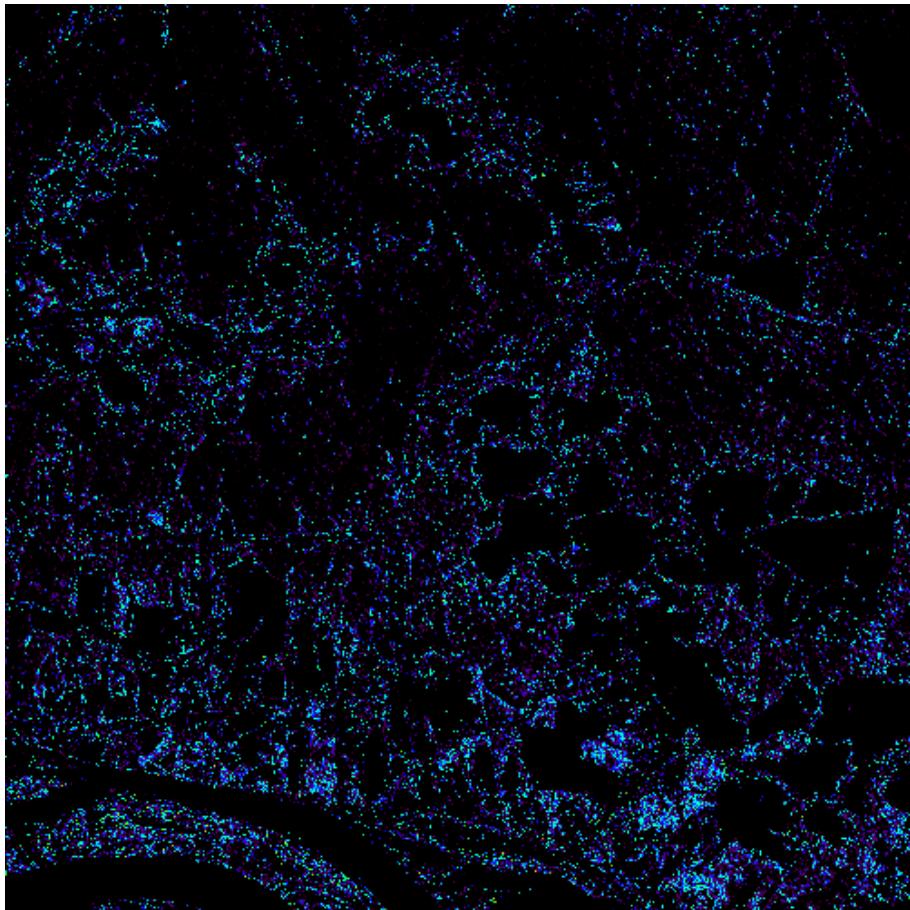
- Floresta (3262 pixels)
- Regeneração (379 pixels)
- Desmatamento (1371 pixels)
- Queimada (392 pixels)
- Água (1186 pixels)



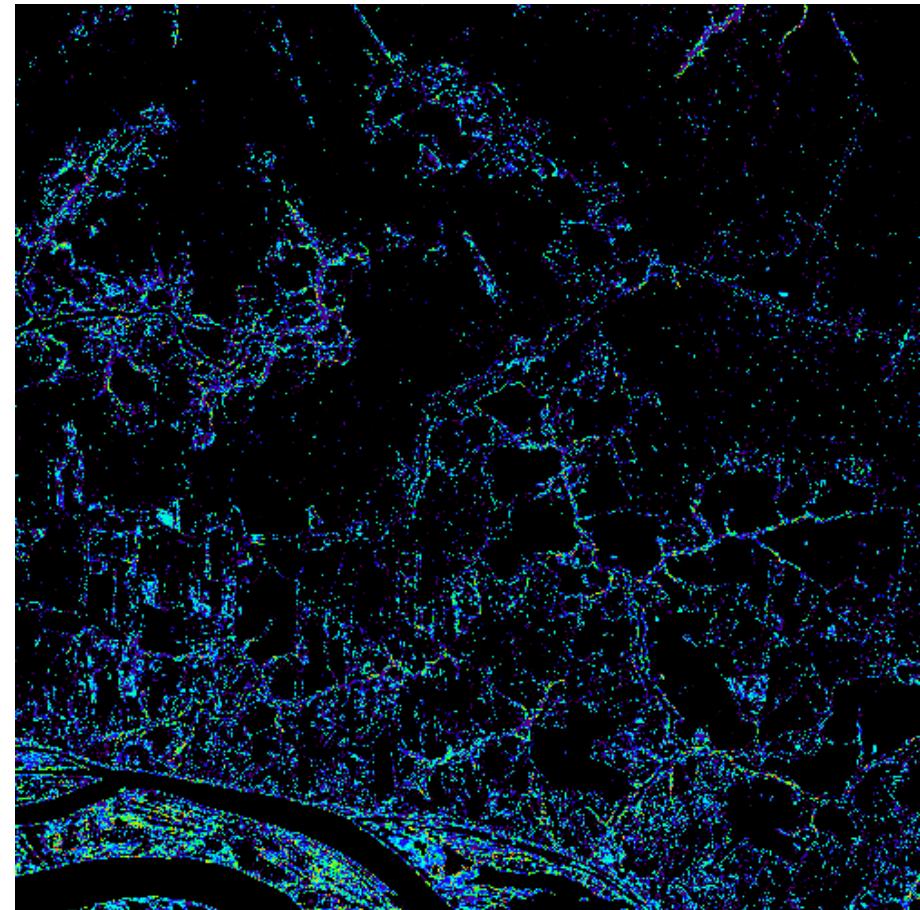
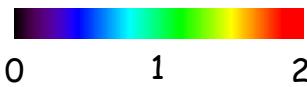
Faz-se a classificação MaxVer escolhendo-se aleatoriamente
apenas 10% dos pontos por classe
Repete-se este procedimento 10.000 vezes

Mapa Final: classe majoritária

Mapas de Incerteza - Exemplo 1 e 2



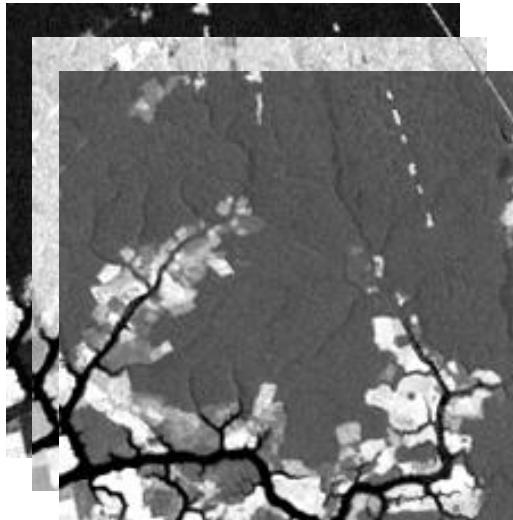
Entropia de Shannon - Exemplo 1



Entropia de Shannon - Exemplo 2

OBS: Áreas com baixa incerteza não implicam em altas exatidões
Mapas de incerteza visam avaliar o desempenho do classificador

Qual impacto do tamanho e estratégia amostral na avaliação do classificador?



Dados

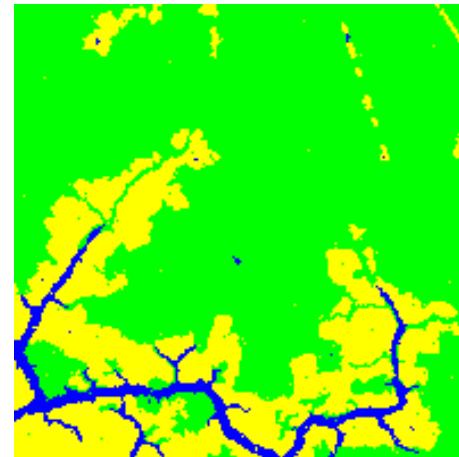


Imagen Classificada

Como avaliar o classificador?

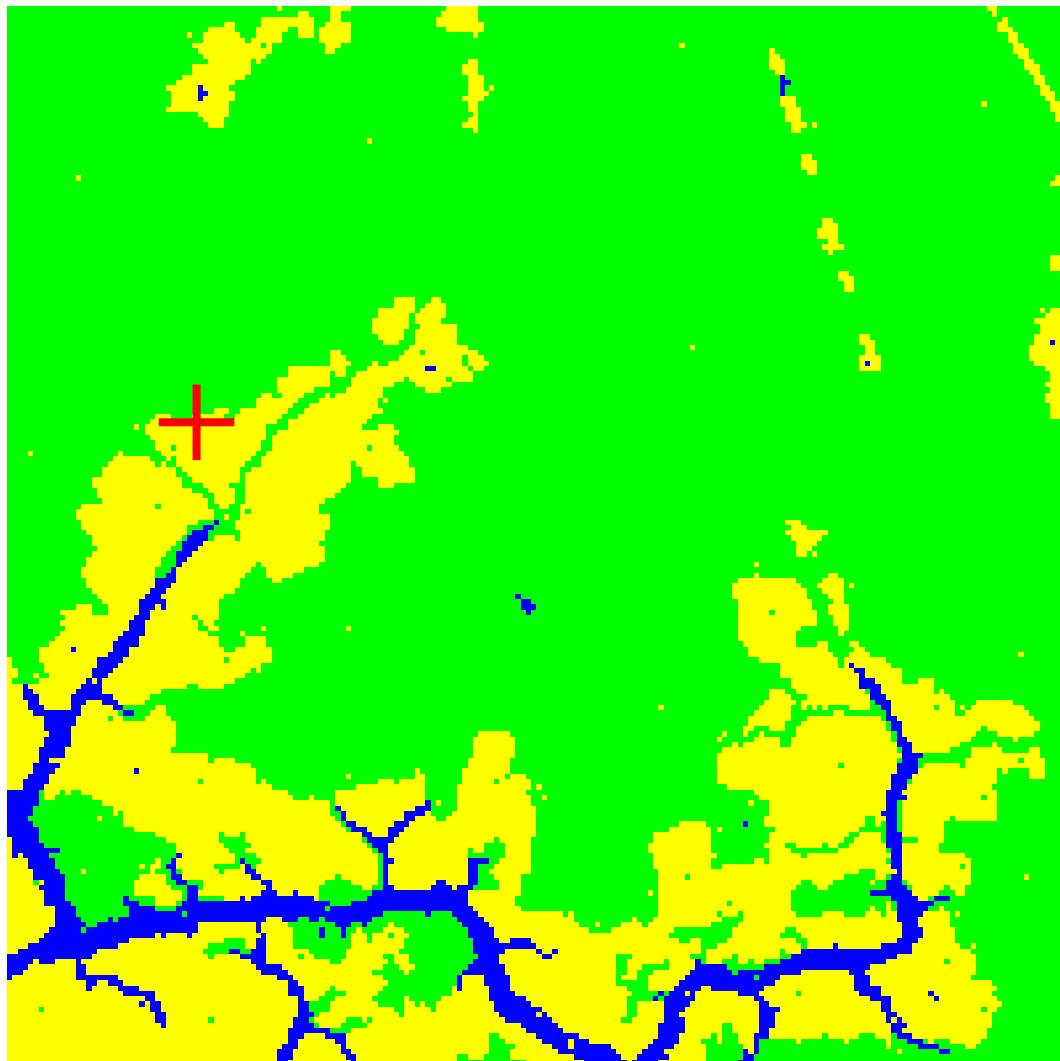
- não se conhece a verdade (referência)
- não se tem ideia de onde estão os erros

Solução: selecionar um conjunto de pontos para verificação!

- quantos pontos selecionar?
- qual estratégia utilizar? aleatória? estratificada?

Qual índice utilizar? exatidão global? erro do usuário?

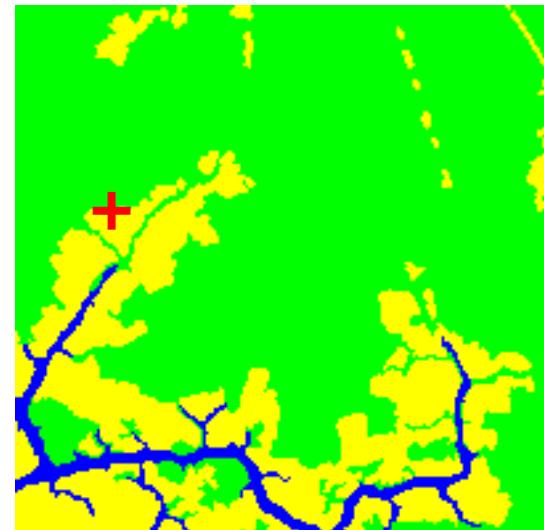
Vamos Simular?



O que podemos esperar

- quanto mais amostras melhor!
- se for avaliar índices globais, abordagens que preservem a proporcionalidade entre classes, devem apresentar melhores resultados

Para isso, vamos supor que "existe" uma verdade que pode ser consultada



Experimentos

Vamos testar 3 tamanhos de amostras:

- 100 pontos (econômica)
- 300 pontos (indicada por Congalton e Olofsson)
- 1000 pontos (exagerada)

Vamos testar 3 abordagens de seleção:

- totalmente aleatória
- estratificada com amostras proporcionais
- estratificada com amostras de mesmo tamanho

Importante:

- como não conhecemos a verdade, iremos usar a imagem classificada para estratificação
- dúvida: será que a proporção de classes na imagem classificada reflete a verdadeira proporção da verdade?

Vamos avaliar:

- exatidão global
- erros do produtor
- erros do usuário

O processo de amostragem e avaliação será repetido 1000 vezes (poderia ser mais!)

Valores de Referência

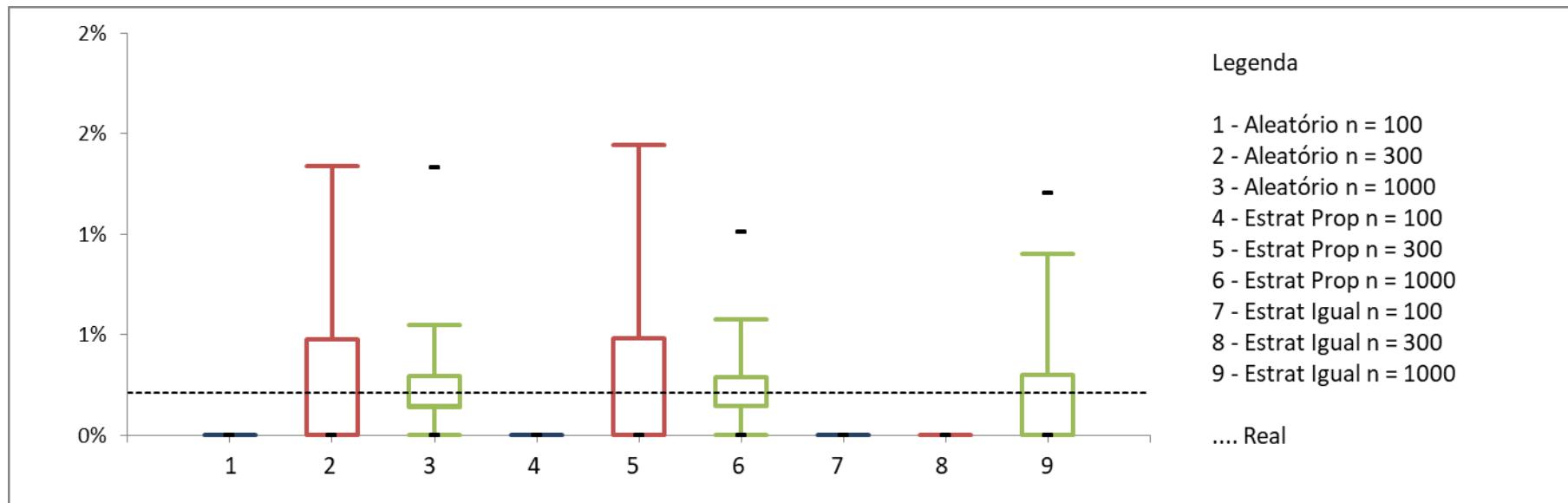
Ao cruzar a imagem classificada com a referência, obtemos os seguintes resultados:

		Referência			} proporção observada
		Floresta	Desmat	Agua	
Class	Floresta	69,20%	0,13%	0,02%	
	Desmat	0,16%	26,19%	0,06%	
	Agua	0,10%	0,07%	4,08%	
		69,46%	26,39%	4,16%	100,00%
} proporção real					

ExatGlobal	Erro do Produtor			Erro do Usuário		
	Floresta	Desmat	Agua	Floresta	Desmat	Agua
99,47%	0,37%	0,75%	1,80%	0,21%	0,80%	3,94%

Erros do Usuário

Floresta



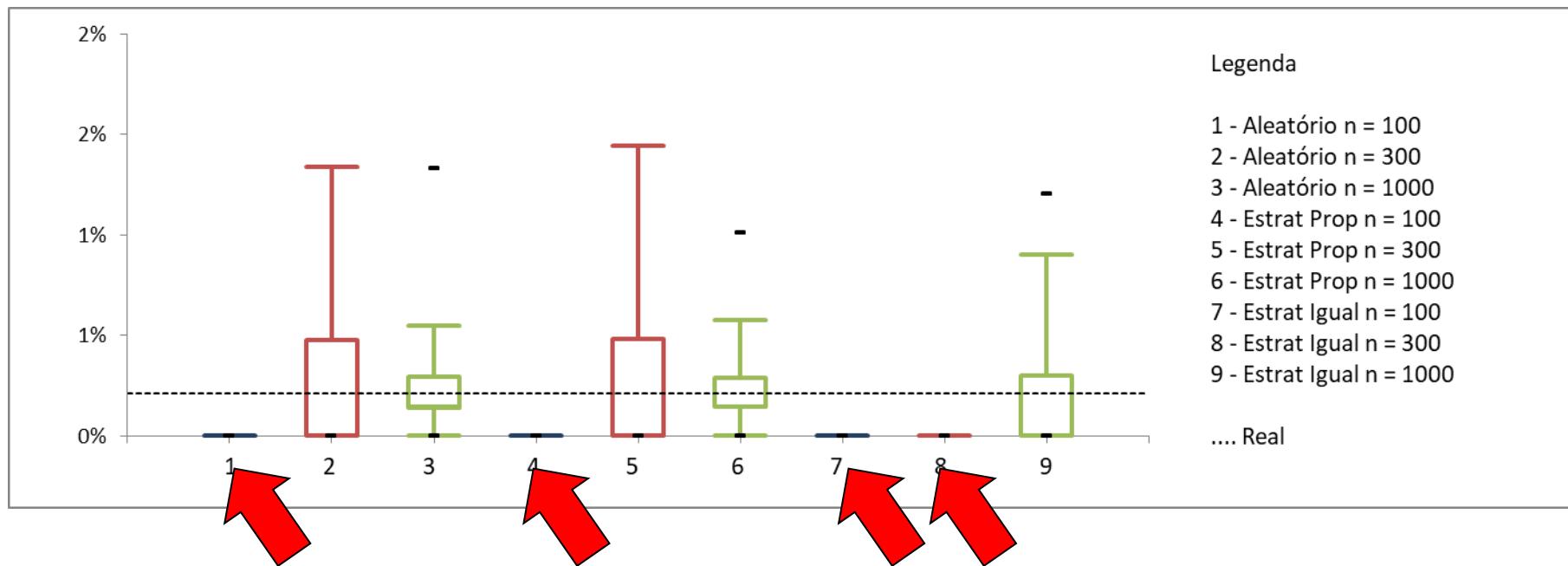
Na estratégia estratificada as amostras são divididas entre as classes:

	100		300		1000	
	Prop	Igual	Prop	Igual	Prop	Igual
Floresta	69	33	208	100	694	333
Desmat	26	33	79	100	264	333
Agua	4	33	13	100	43	333

Mal representada na abordagem proporcional

Erros do Usuário

Floresta

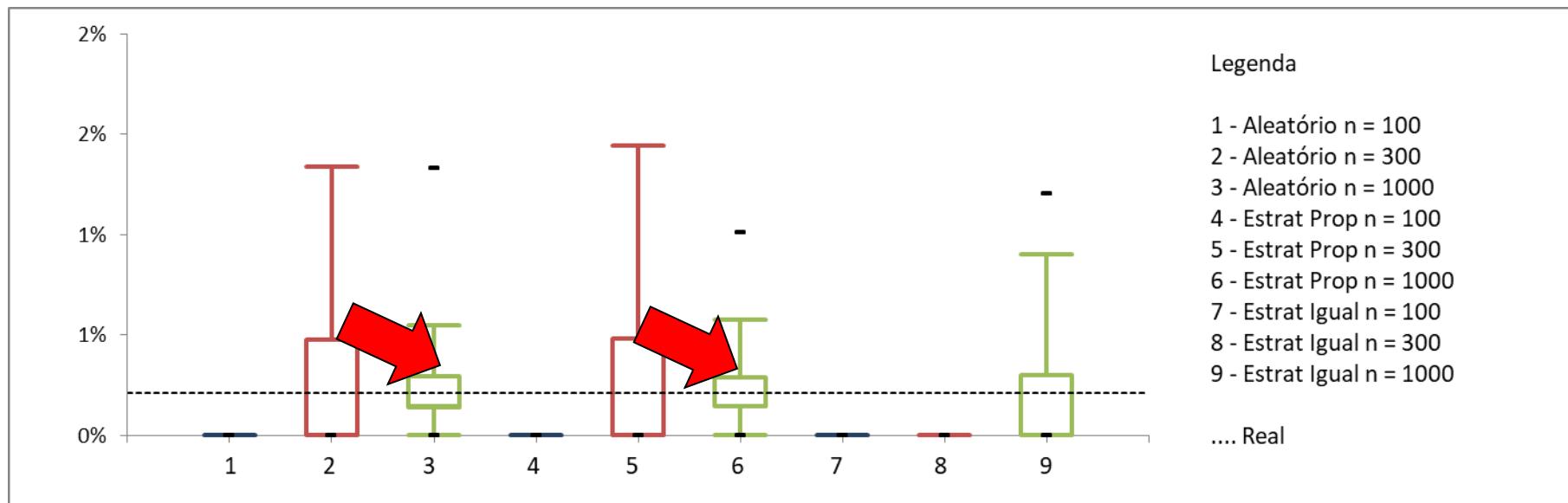


Com 100 amostras apenas, não foi possível avaliar o erro para classe floresta

Fica bem complicado "acertar" o erro numa classe tão abrangente (~70% da área total)

Erros do Usuário

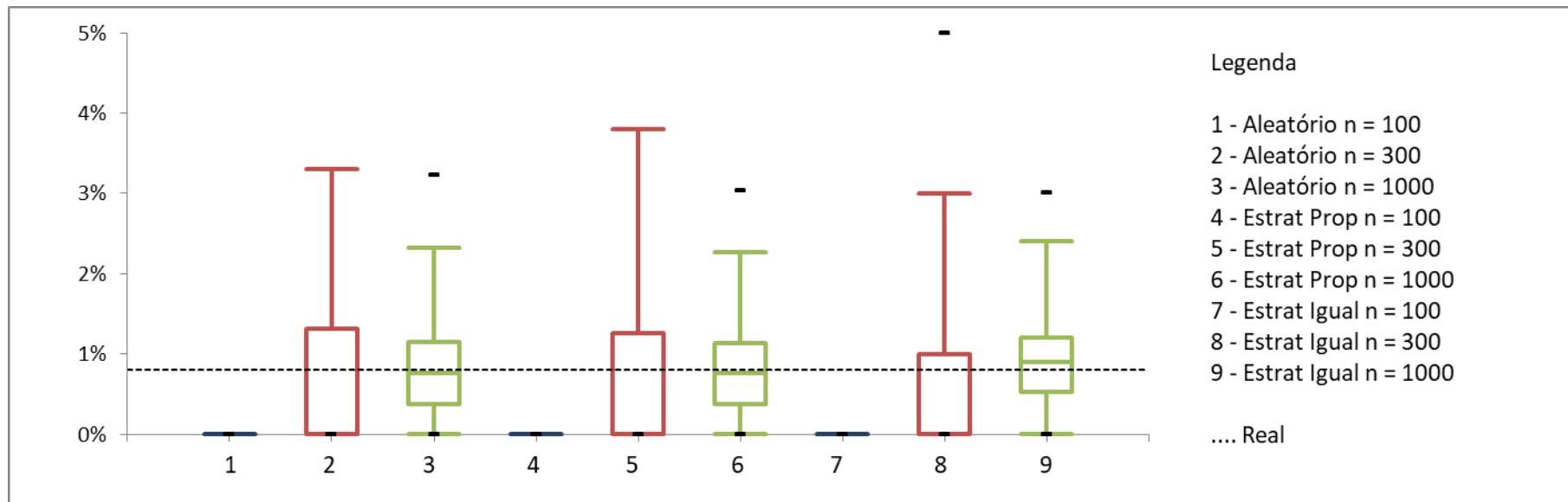
Floresta



Com o aumento do tamanho da amostra, diminui-se a incerteza e melhora-se a precisão na estimativa (mediana próximo ao real), mas ainda assim pode-se observar valores bastante atípicos (distribuição bastante assimétrica para o zero)

Erros do Usuário

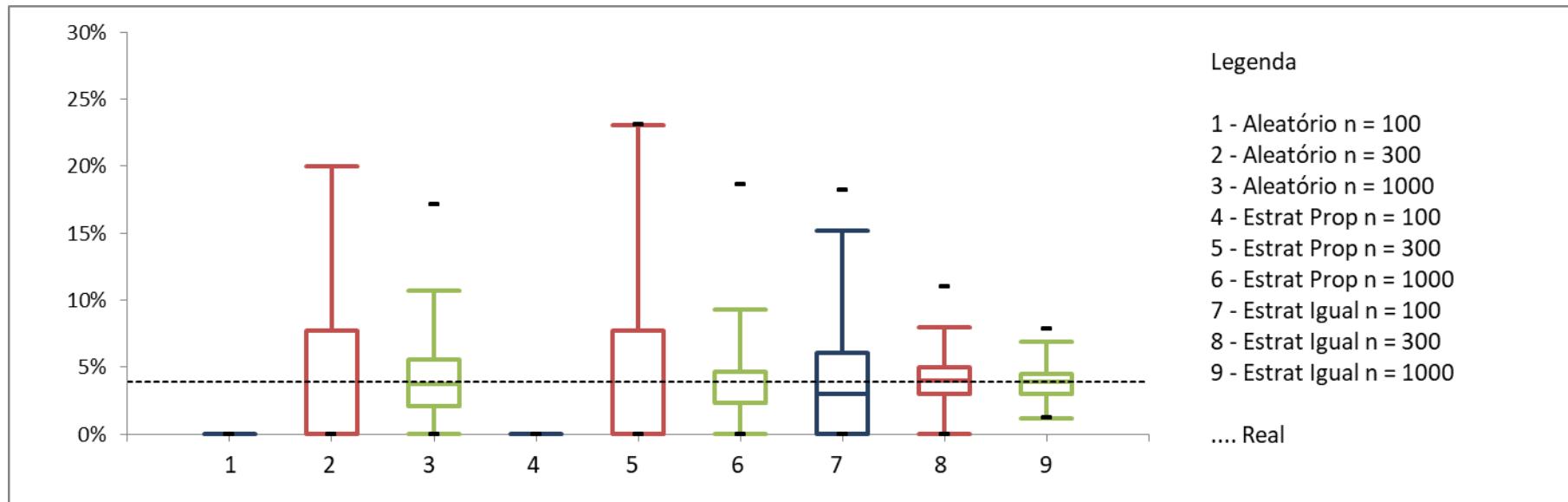
Desmat



Mesmo comportamento para a classe Desmat

Erros do Usuário

Água

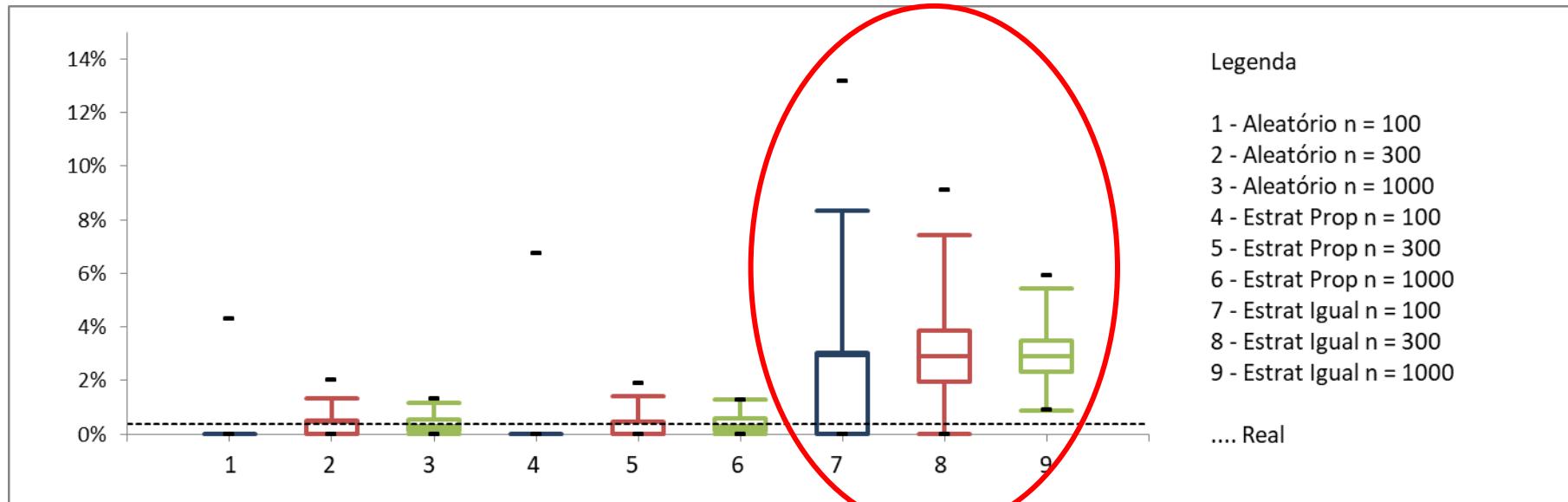


Para classe água, os resultados são bastante variados

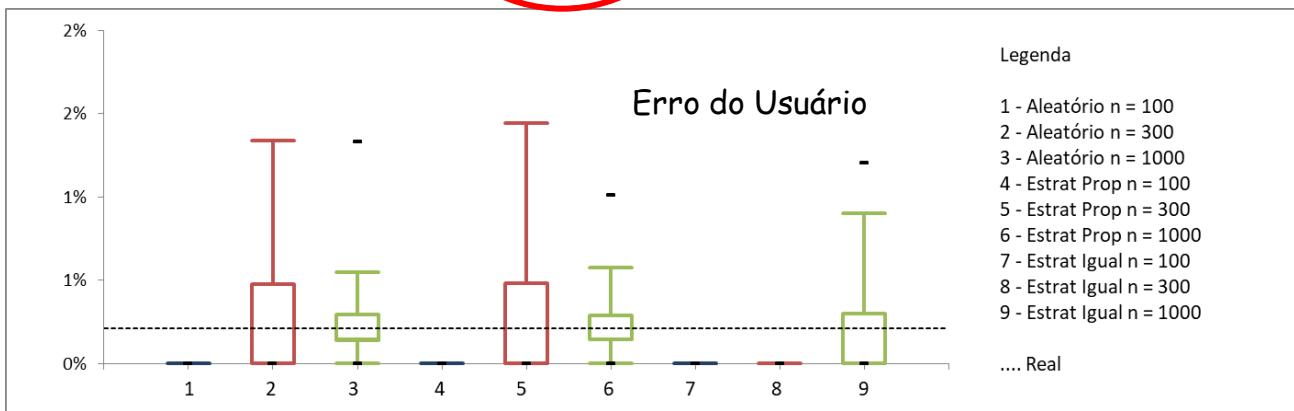
As estimativas são mais precisas quando usado a abordagem estratificada com mesmo número de amostras por classe

Erros do Produtor

Floresta

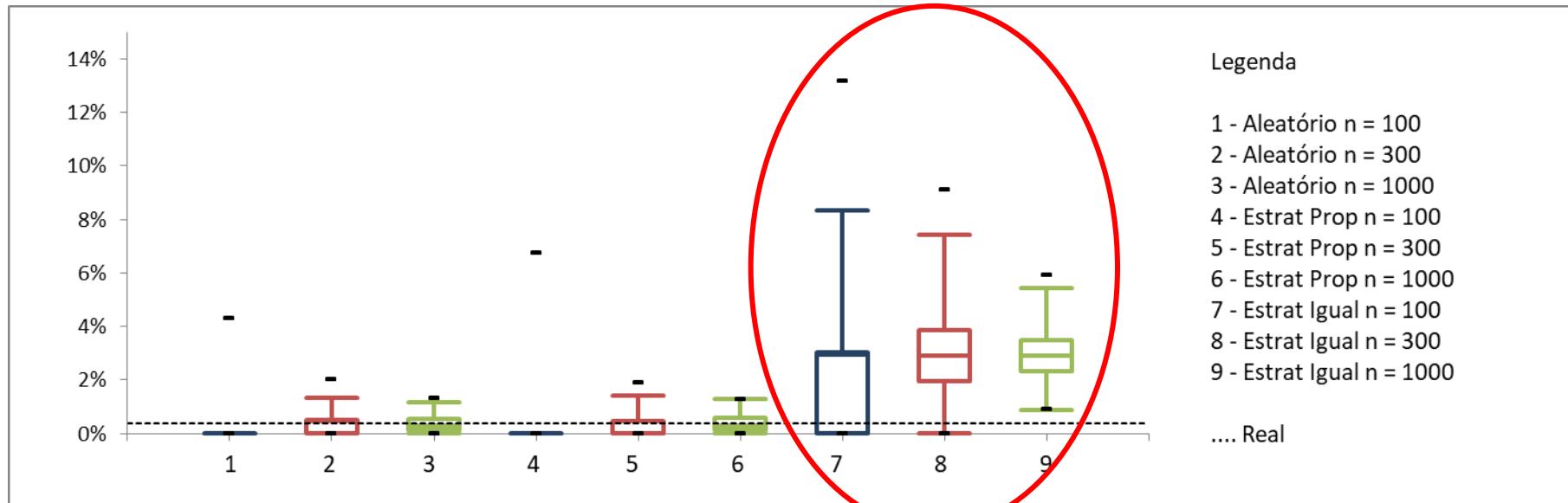


Como explicar isso?



Erros do Produtor

Floresta



Exemplo de simulação:

		Referência		
		Floresta	Desmat	Agua
Class	Floresta	332	1	0
	Desmat	4	327	2
	Agua	12	5	316
		348	333	318
		999		

Erro do usuário real da classe água é bem maior que o erro do produtor real!

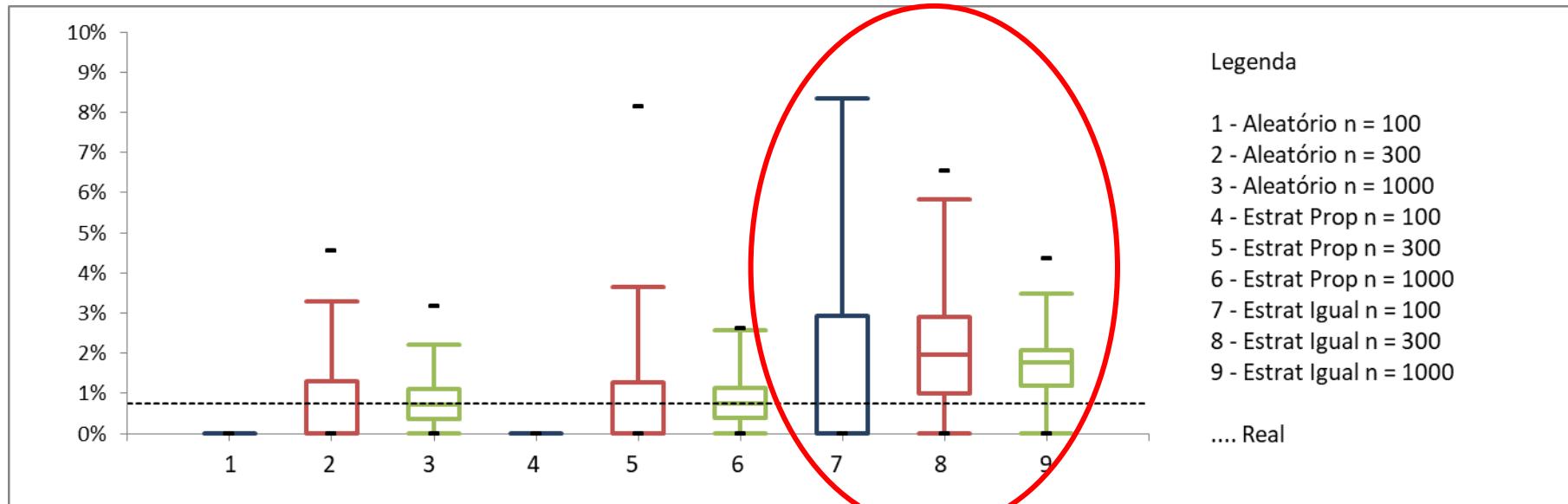
Erro Produtor = 1,80%

Erro Usuário = 3,94%

Ao amostrar mais esta classe, amostrou-se também mais erros dessa classe.

Erros do Produtor

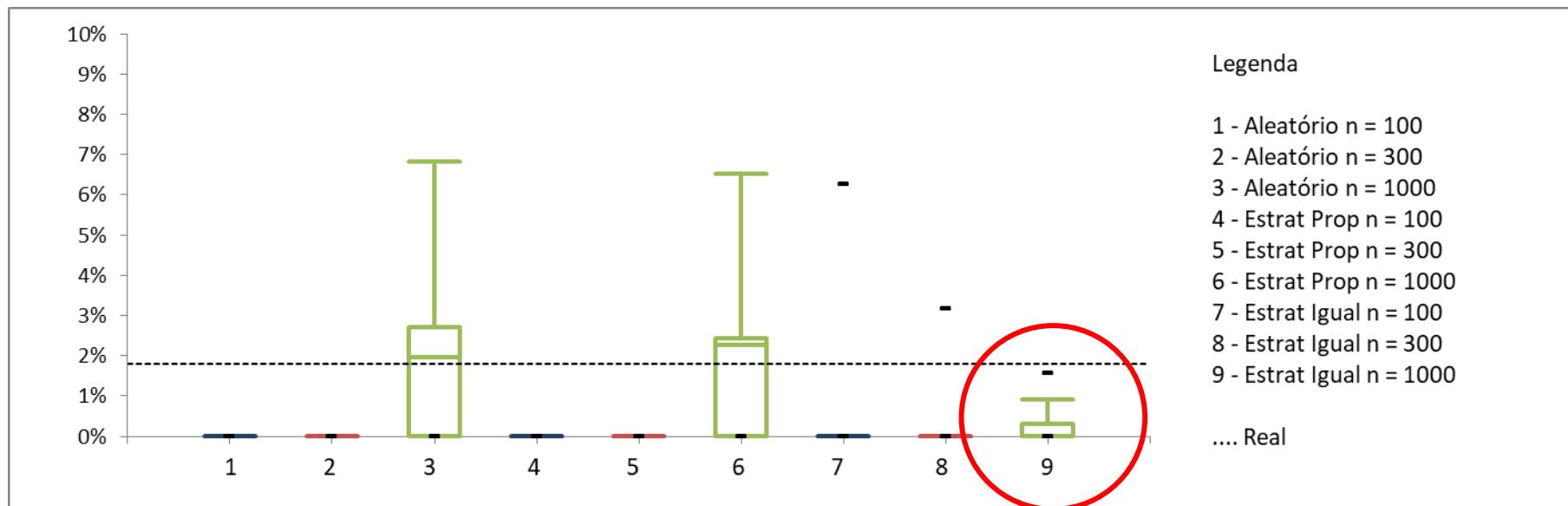
Desmat



Também foi prejudicado na abordagem estratificada com mesmo tamanho de amostra para todas as classes!

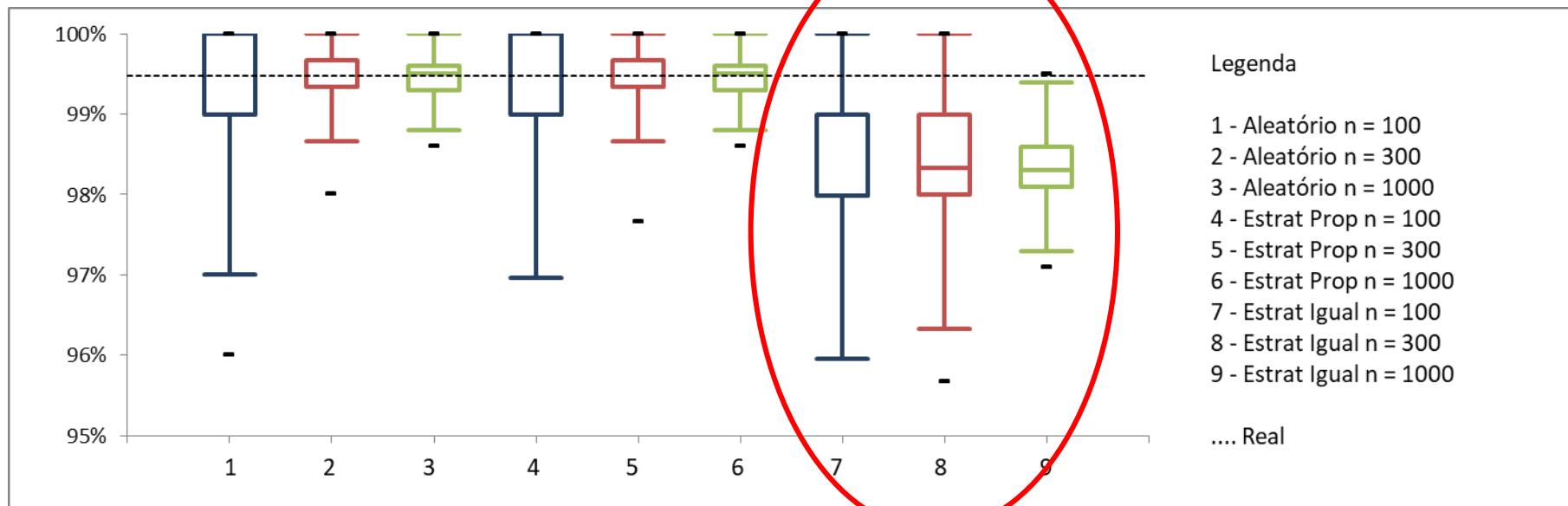
Erros do Produtor

Agua



Diminuiu a variabilidade mas também subestimou o erro do produtor para essa classe

Exatidão Global

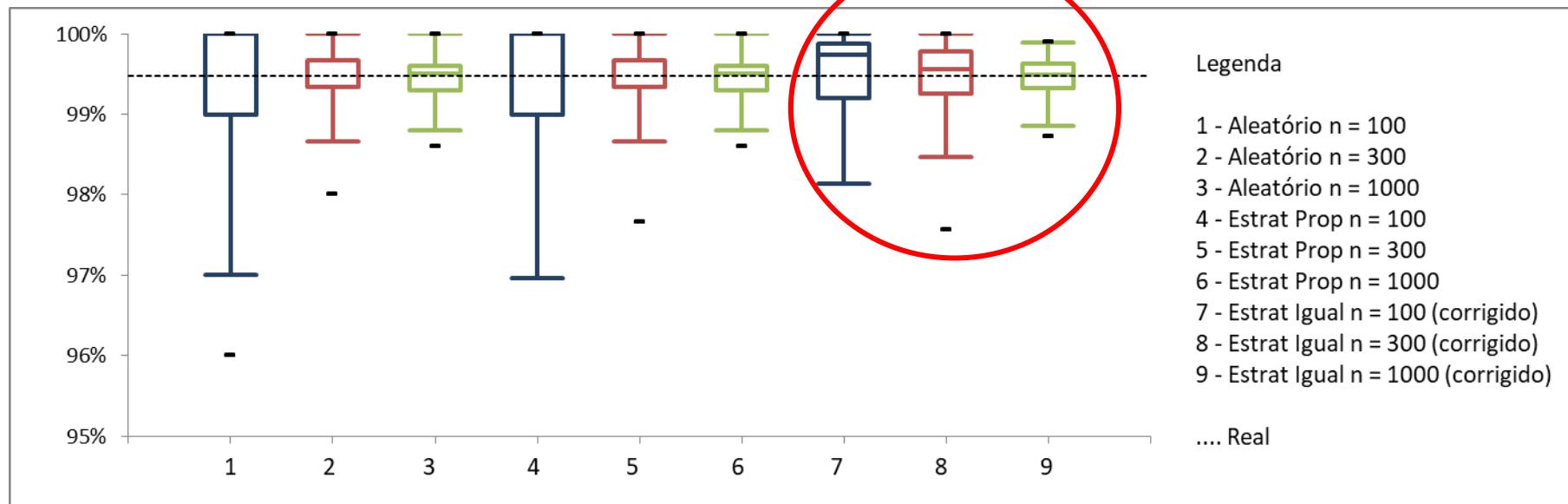


Aparentemente, a estratificação usando mesmo tamanho de amostra para todas as classes prejudica a estimativa da exatidão global

Nesse caso, perde-se a proporcionalidade de cada classe (pesos iguais para todas as classes)

E se corrigíssemos pela proporção das classes observadas na imagem classificada?

Exatidão Global

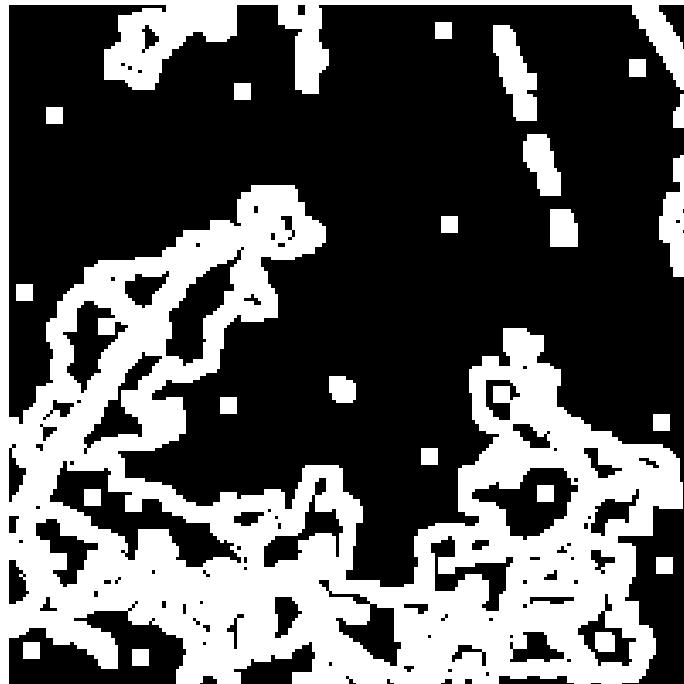
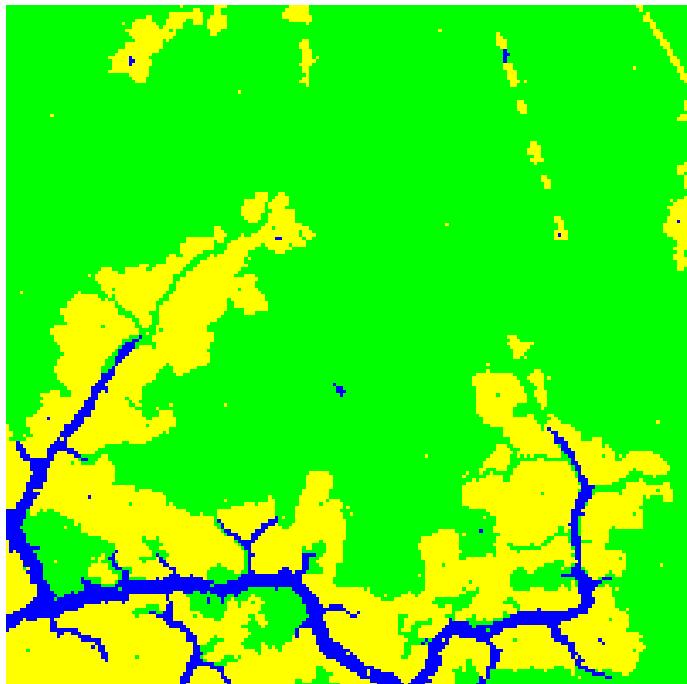


Houve uma melhora significativa nas estimativas

Mas e se soubéssemos de algum padrão espacial da ocorrência de erros?

Por exemplo, se tivéssemos evidências para acreditar que o classificador erra mais próximo às bordas?

Exatidão Global



As amostras foram divididas em 2 partes iguais (borda e não borda) e distribuídas proporcionalmente para cada classe

Exatidão Global

