

Aprendizado de Máquina

Aula 3.3 - Classificação multi-classe


Adriano Rivolli

rivolli@utfpr.edu.br

Especialização em Inteligência Artificial


Universidade Tecnológica Federal do Paraná (UTFPR)
Câmpus Cornélio Procopio
Departamento de Computação

Conteúdo

- 
- 1 Decomposição de problemas
 - 2 Medidas de avaliação

Decomposição de problemas


Problemas multi-classe

- 
- Quando o atributo alvo possui mais de duas classes
 - Alguns algoritmos não suportam nativamente
 - ▶ É possível transformar um problema multiclasse em N problemas binários
 - As medidas de avaliação também precisam de ajuste

Estratégias

- **One-versus-all** (OVA) ou **One-verus-rest** (OVR)
 - ▶ Um contra todos
 - ▶ Um conjunto de dados para cada classe
- **One-versus-one** (OVO)
 - ▶ Um contra um (um contra outro)
 - ▶ Um conjunto de dados para cada par de classes

Um contra todos

- 
- Um conjunto de dados para cada classe
 - ▶ As instâncias que possuem a classe serão positivas
 - ▶ As instâncias das demais classes serão negativas
 - Aumenta a taxa de desbalanceamento
 - Induz um modelo para cada conjunto binário
 - ▶ A maior probabilidade define a classe da instância

Exemplo de OVA

Multi-classe

Attr	Classe
X ₁	A
X ₂	A
X ₃	B
X ₄	B
X ₅	B
X ₆	C
X ₇	C

Classe A

Attr	Classe
X ₁	1
X ₂	1
X ₃	0
X ₄	0
X ₅	0
X ₆	0
X ₇	0

Classe B

Attr	Classe
X ₁	0
X ₂	0
X ₃	1
X ₄	1
X ₅	1
X ₆	0
X ₇	0

Classe C

Attr	Classe
X ₁	0
X ₂	0
X ₃	0
X ₄	0
X ₅	0
X ₆	1
X ₇	1

Um contra um

- Um conjunto de dados para cada par de classes
 - ▶ Ao todo são gerados $\frac{q(q-1)}{2}$ conjuntos
 - ▶ Uma das classes assume a classe positiva e a outra a classe negativa
 - ▶ Os demais exemplos são descartados
- Induz um modelo para cada conjunto binário
 - ▶ Cada modelo vota em uma classe (classe predita)
 - ▶ A classe que receber o maior número de votos é predita

Exemplo de OVO

Multi-classe

Attr	Classe
X ₁	A
X ₂	A
X ₃	B
X ₄	B
X ₅	B
X ₆	C
X ₇	C

Classe A x B

Attr	Classe
X ₁	1
X ₂	1
X ₃	0
X ₄	0
X ₅	0

Classe B x C

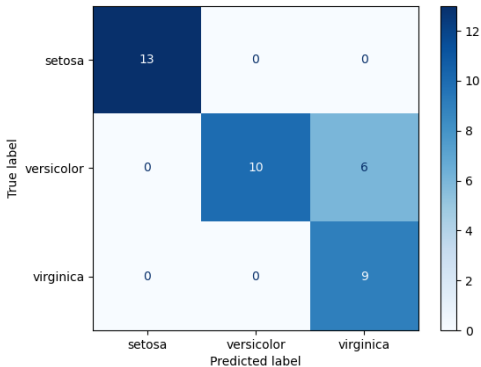
Attr	Classe
X ₃	1
X ₄	1
X ₅	1
X ₆	0
X ₇	0

Classe A x C

Attr	Classe
X ₁	1
X ₂	1
X ₆	0
X ₇	0

Medidas de avaliação

Matriz de confusão



Fonte: https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html

Abordagens

■ *Micro-averaged*

- ▶ Usa uma única matriz de confusão (totalizando todas as classes)
- ▶ Trata cada instância igualmente

■ *Macro-averaged*

- ▶ Usa uma matriz de confusão para cada classe e calcula a média
- ▶ Trata cada classe igualmente

■ *Weighted-averaged*

- ▶ Abordagem similar a macro, porém usa a distribuição das classes como pesos

Acurácia

■ Não há diferença entre a acurácia micro e macro

■ Micro:

$$Acc = \frac{\sum_{j=0}^q TP_j + \sum_{j=0}^q TN_j}{\sum_{j=0}^q TP_j + \sum_{j=0}^q TN_j + \sum_{j=0}^q FP_j + \sum_{j=0}^q FN_j}$$

■ Macro:

$$Acc = \frac{1}{q} \sum_{j=0}^q Acc_j$$

Precisão

■ Micro:

$$Prec = \frac{\sum_{j=0}^q TP_j}{\sum_{j=0}^q TP_j + \sum_{j=0}^q FP_j}$$

■ Macro:

$$Prec = \frac{1}{q} \sum_{j=0}^q Prec_j$$

Revocação

■ Micro:

$$Rec = \frac{\sum_{j=0}^q TP_j}{\sum_{j=0}^q TP_j + \sum_{j=0}^q FN_j}$$

■ Macro:

$$Rec = \frac{1}{q} \sum_{j=0}^q Rec_j$$

$F1$

■ Micro:

$$F1 = 2 \frac{precision_{micro} \cdot recall_{micro}}{precision_{micro} + recall_{micro}}$$

■ Macro:

$$F1 = \frac{1}{q} \sum_{j=0}^q F1_j$$