

Técnicas de Ensemble

Vinicius Teixeira de Melo

viniciusteix@alu.ufc.br



Agenda

- 1 Ensemble
- 2 Combinando Classificadores
- 3 Bagging
- 4 CIFAR 10

Ensemble

Introdução

- 1 Combinar diferentes classificadores;
- 2 Generalizar o melhor de cada classificador;
- 3 Formar um novo classificador.

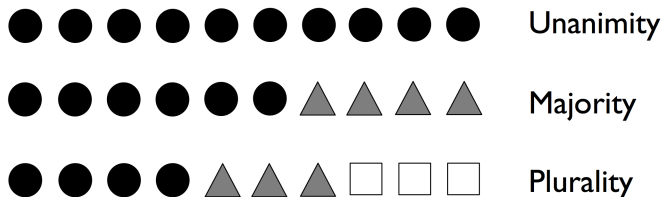


Figura: Tipos de votação

Ensemble

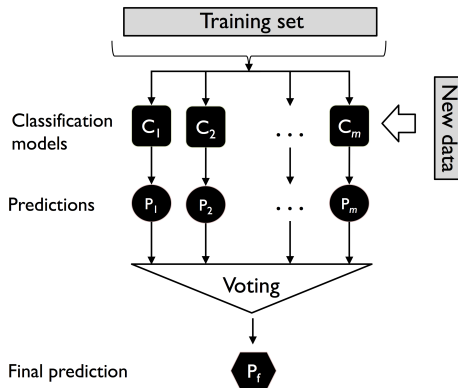


Figura: Abordagem utilizando votação majoritária

Votação simples

$$y = \text{mode}\{C_1(x), C_2(x), \dots, C_m(x)\}$$

Exemplo

Obs: $\text{class1} = -1$ e $\text{class2} = +1$

$$C(x) = \text{sign} \left[\sum_j^m C_j(x) \right] = \begin{cases} 1 & \text{if } \sum_i C_j(x) \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

Combinando Classificadores

Implementação de um classificador por votação simples

- Combinar diferentes classificadores associados a pesos individuais;
- Equilibrar as fraquezas dos classificadores individuais;
- Voto majoritário ponderado;

Voto majoritário ponderado

$$y = \operatorname{argmax}_i \sum_{j=1}^m w_j X_A (C_j(x) = i)$$

Equação simplificada (pesos iguais)

$$y = \operatorname{mode}\{C_1(x), C_2(x), \dots, C_m(x)\}$$

Exemplo

- Predizer a *label* de uma amostra x ;
- $i \in \{0, 1\}$;
- $C_j(j \in \{1, 2, 3\})$;
- $C_1(x) = 0, C_2(x) = 0, C_3(x) = 1$.

Voto majoritário ponderado

$$y = \underset{i}{\operatorname{argmax}} \sum_{j=1}^m w_j X_A (C_j(x) = i)$$
$$= \underset{i}{\operatorname{argmax}} [0.2i_0 + 0.2i_0 + 0.6i_1] = 1$$

Equação simplificada (pesos iguais)

$$y = \operatorname{mode}\{0, 0, 1\} = 0$$

Classificador individual retorna uma probabilidade

$$y = \operatorname{argmax}_i \sum_{j=1}^m w_j p_{ij}$$

Exemplo

$C_1(x) \rightarrow [0.9, 0.1]$, $C_2(x) \rightarrow [0.8, 0.2]$, $C_3(x) \rightarrow [0.4, 0.6]$

$$p(i_0|x) = 0.2 \times 0.9 + 0.2 \times 0.8 + 0.6 \times 0.4 = 0.58$$

$$p(i_1|x) = 0.2 \times 0.1 + 0.2 \times 0.2 + 0.6 \times 0.6 = 0.42$$

$$y = \operatorname{argmax}_i [p(i_0|x), p(i_1|x)] = 0$$

Notebook

<https://github.com/viniciusteix/topics-3/blob/master/seminario>

Combinando Classificadores

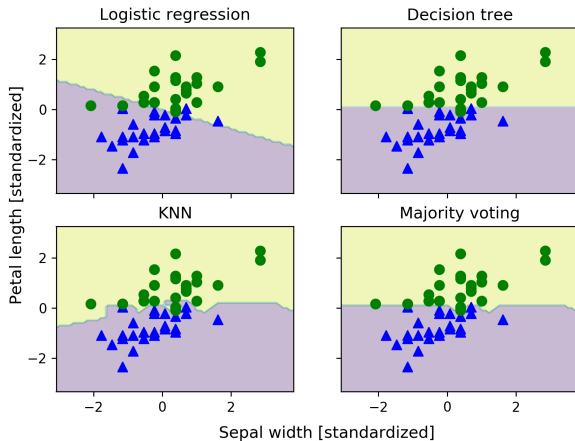


Figura: Separação das classes

Bagging

Introdução

- Relacionada a técnica de Classificação por votação majoritária;
- Utilizada amostras aleatórias em cada classificador;
- Árvores de decisão são, tipicamente, utilizadas como classificadores base.

Bagging

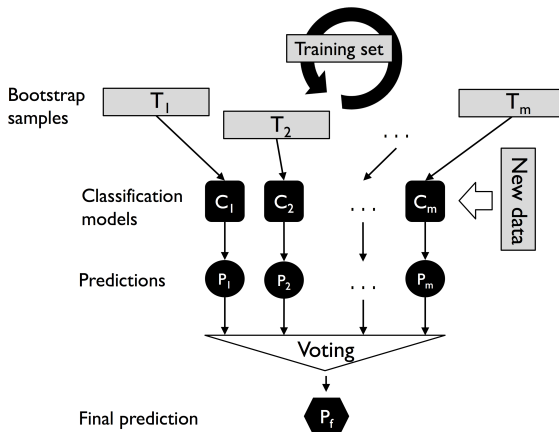
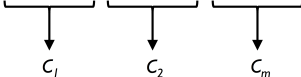


Figura: Abordagem utilizando Bagging

Bagging

Sample indices	Bagging round 1	Bagging round 2	...
1	2	7	...
2	2	3	...
3	1	2	...
4	3	1	...
5	7	1	...
6	2	7	...
7	4	7	...



C_1 C_2 C_m

Figura: Amostras aleatórias para o Bagging

Notebook

<https://github.com/viniciusteix/topics-3/blob/master/seminario>

CIFAR 10

CIFAR 10

airplane



automobile



bird



cat



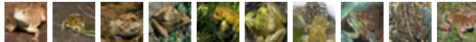
deer



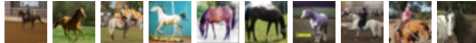
dog



frog



horse



ship



truck



Figura: Classes do CIFAR 10

Notebook

<https://github.com/viniciusteix/topics-3/blob/master/seminario>

Obrigado