

Aprendizagem de Máquina

Alceu S. Britto

Programa de Pós-Graduação em Informática
Pontifícia Universidade Católica do Paraná (PUCPR)

Ensembles e Combinação de Classificadores

Filosofia

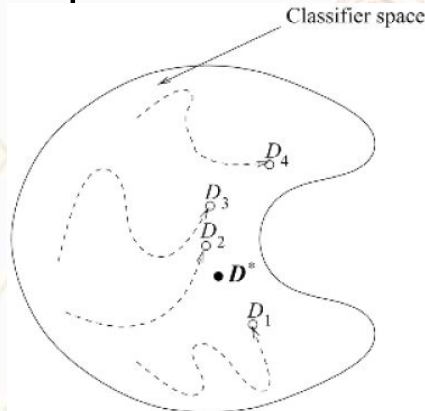
- O objetivo da combinação de classificadores é:

Buscar uma classificação mais precisa pagando o preço de uma maior complexidade.

- Ao invés de buscar o melhor conjunto de características e o melhor classificador, buscamos agora **o melhor conjunto de classificadores** e então **o melhor método de combinação**.

Motivação

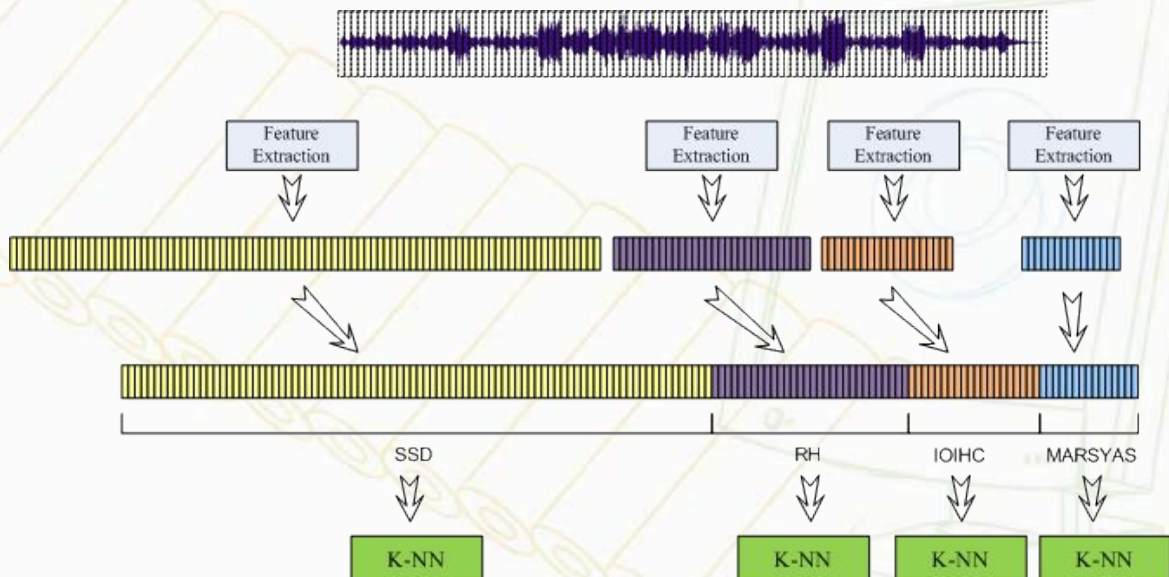
- Suponha que temos um conjunto de dados Z rotulado e diversos classificadores (D_1, D_2, \dots) com um bom desempenho em Z .



- Podemos pegar um único classificador como solução, correndo o risco de tomar uma má decisão para o problema.

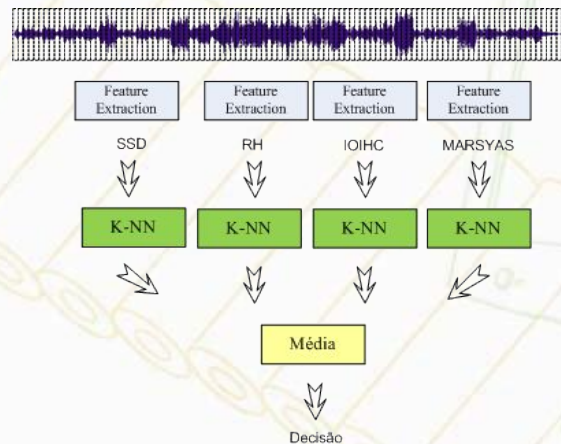
Motivação

- Exemplo: suponha que executamos um classificador 1-NN para 4 subconjuntos diferentes de características, obtendo como consequência 4 classificadores com erro zero no conjunto de treinamento.



Motivação

- Apesar destes classificadores terem todos o mesmo desempenho sobre o conjunto de treinamento
 - eles podem ter diferentes desempenhos na generalização!!!
- Ao invés de pegar somente um classificador, uma opção mais segura seria usá-los todos e fazer a média de suas saídas.

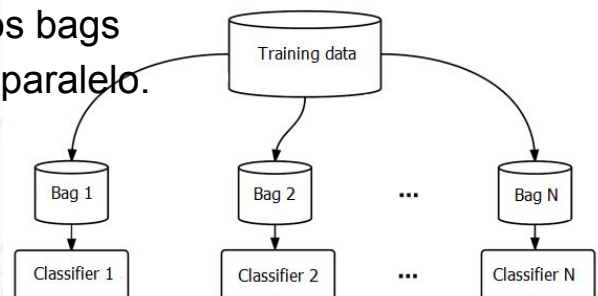


Motivação

- Mas este conjunto de classificadores pode não ser melhor do que o melhor classificador individual!
- Mas diminuirá ou eliminará o risco de escolhermos um classificador individual inadequado!

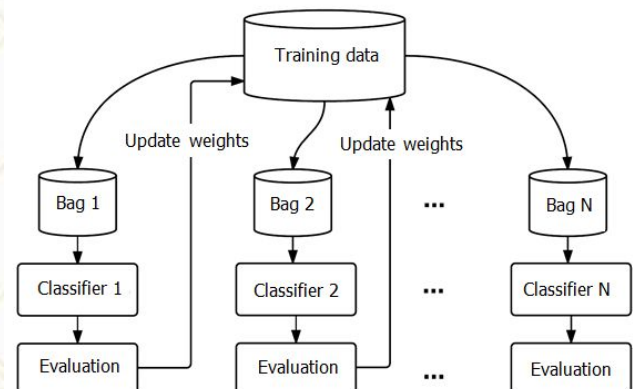
Ensembles (Comitês)

- Bagging (Breiman, 1996) or Bootstrapped Aggregation
 - Gera ensemble de classificadores homogêneos treinados a partir de amostragem randômica com repetições.
 - Qualquer instância tem a mesma probabilidade de aparecer em um novo bag (amostra de dados de treinamento).
 - Principais características
 - escolhe-se o estimador de base
 - define-se a quantidade de estimadores
 - pode-se definir o tamanho dos bags
 - classificadores treinados em paralelo.



Ensembles (Comitês)

- Boosting (R.E. Schapire, et al., 1997)
 - Gera ensemble de classificadores homogêneos treinados a partir de amostragem randômica (com repetições) em dados ponderados
 - Instância difíceis têm maior probabilidade de aparecer em um novo bag (amostra de dados com pesos)
 - Principais características:
 - escolhe-se o estimador de base
 - definir-se a quantidade de estimadores.
 - treinamento acontece de forma sequencial.



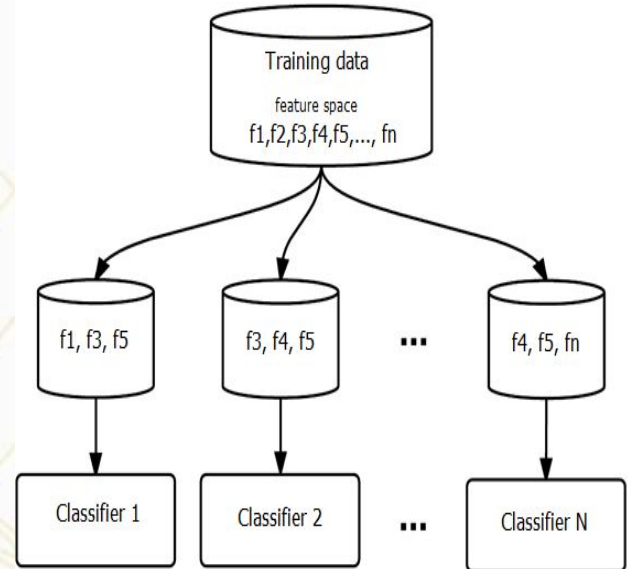
Ensembles (Comitês)

- Random Subspace (T.K. Ho, 1998)

- Features são amostradas randomicamente.

- Principais Características

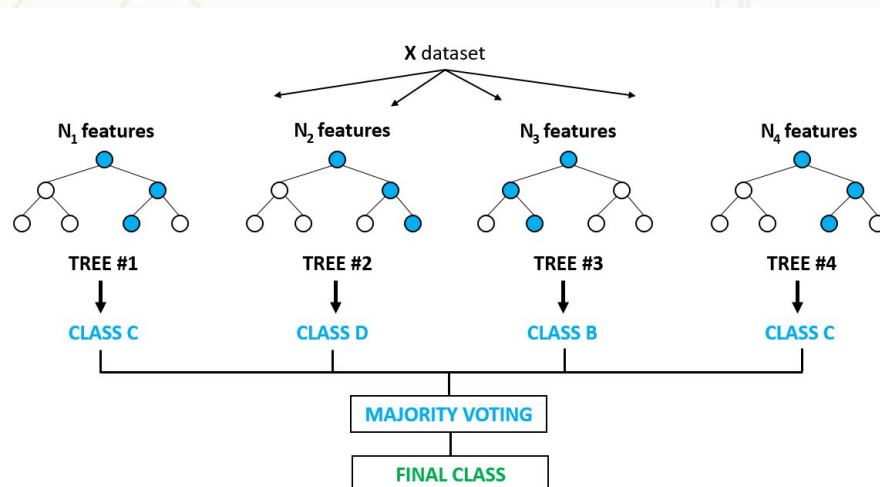
- escolhe-se o estimador de base
 - define-se a % de features a ser randomicamente selecionada para o treinamento dos membros do ensemble.



Ensembles (Comitês)

- Floresta Aleatória (Random Forest)

- Cria uma combinação (*ensemble*) de árvores de decisão, na maioria dos casos treinados com o método *bagging* e definindo o melhor atributo para cada nó a partir de um subconjunto de atributos também definido aleatoriamente.



Ensembles (Comitês)

- XGBoost (Extreme Gradient Boosting)
 - Cria um *ensemble* de árvores.
 - Utiliza medida de similarity como forma de calcular ganho de informação.
 - Descida do gradiente é utilizada para definir a importância de cada membro do ensemble.
 - Função objetivo considera a soma de uma função de perda avaliada sobre a base de treinamento e uma função de regularização que penaliza árvores complexas (prioriza-se árvores simples). Na equação f_j é a predição da j_{th} árvore.

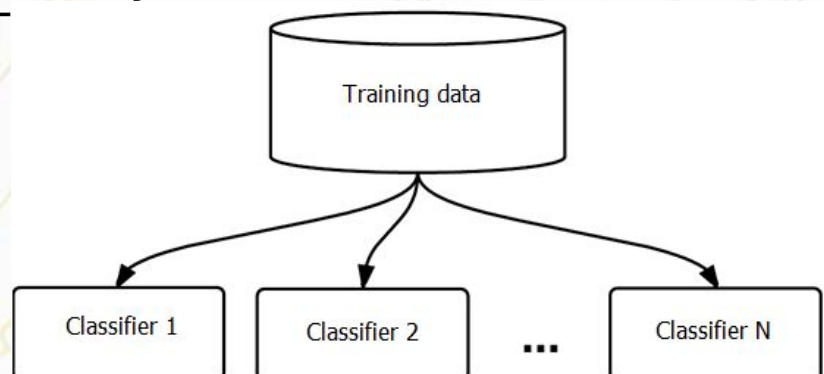
Função Objetivo

$$obj(\theta) = \sum_i^n l(y_i - \hat{y}_i) + \sum_{j=1}^J \Omega(f_j)$$

\uparrow função de perda (loss) \uparrow regularizador

Ensembles (Comitês)

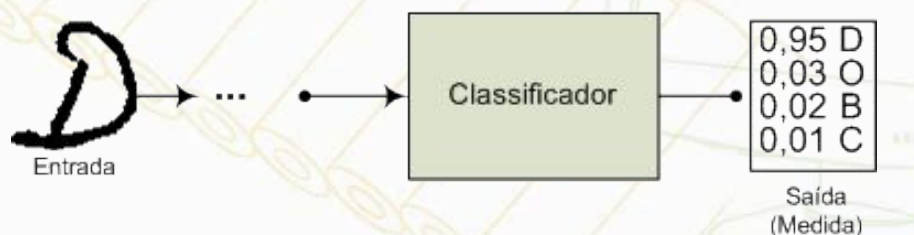
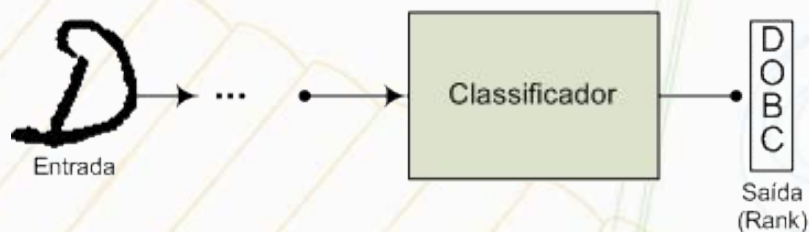
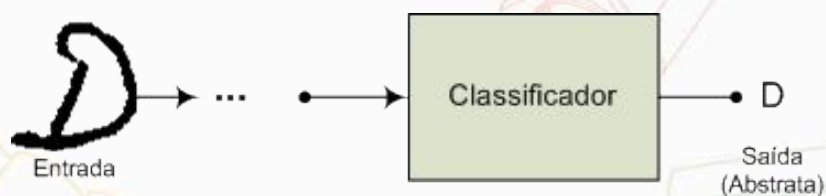
- Ensembles heterogêneos
 - Diferente das abordagens homogêneas que buscam diversidade via manipulação dos dados de treinamento, aqui a diversidade é explorada por meio da combinação de diferentes conceitos.
- Por exemplo: combinação de estimadores baseados em KNN, Decision T



Tipos de Saídas de Classificadores

- As maneiras possíveis de combinar as saídas de L classificadores em um *ensemble*, depende da informação obtida dos membros individuais.
- Podemos distinguir três (3) tipos de saídas de classificadores:
 - Tipo 1: Nível Abstrato: Cada classificador produz um rótulo
 - Tipo 2: Nível de Rank: Cada classificador produz uma lista ordenada de alternativas de acordo com a plausibilidade.
 - Tipo 3: Nível de Medidas: Cada classificador produz um vetor de c dimensões onde cada valor representa o suporte para o vetor.

Tipos de Saídas de Classificadores



Voto da Maioria (Majority Vote)

Unanimity
(all agree)



Simple majority
(50%+1)



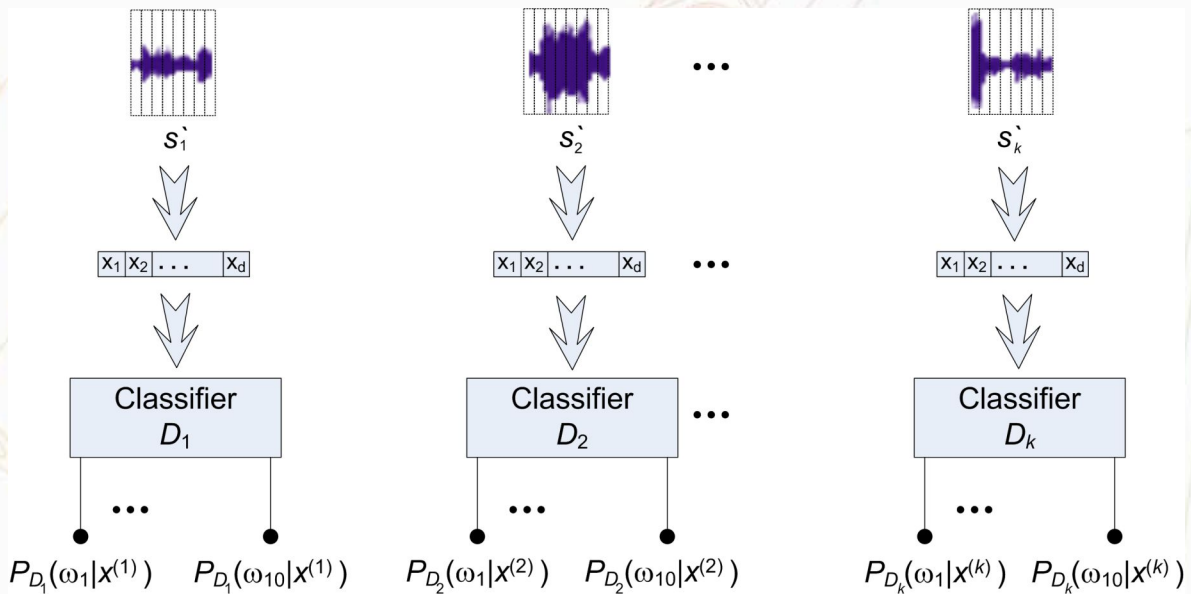
Plurality
(most votes)



Estratégias de Combinação

- Assumimos um vetor d -dimensional de características, x .
- Assumimos que existem c classes possíveis rotuladas w_1 a w_c organizado como um conjunto de rótulos.
- Assumimos um conjunto de k classificadores $D_1 \dots D_k$ onde cada classificador D_i produz na saída $[P_{D_i}(w_1|x), P_{D_i}(w_2|x), \dots, P_{D_i}(w_j|x)]$, onde $P_{D_i}(w_j|x)$ representa o suporte para que hipótese de que o vetor x seja da classe w_j .

Estratégias de Combinação



Votação da Maioria

- Regra de decisão simples, onde somente os rótulos atribuídos pelo classificador são levados em conta e aquele que tiver mais votos é o vencedor.

$$\hat{\omega} = \max_{i \in [1, k]} \text{count} \left[\arg \max_{\omega \in \Omega} P_{D_i}(\omega | x) \right]$$

- onde i indica o índice do classificador e P_{D_i} indica o nível de confiança fornecido na saída do classificador D_i

Max

- Regra de decisão simples, onde a classe com o nível de confiança mais elevado é a vencedora, pouco importando o classificador.

$$\hat{\omega} = \arg \max_{\substack{i \in [1, k] \\ \omega \in \Omega}} P_{D_i}(\omega | x)$$

Soma

- A regra da soma é baseada no somatório dos níveis de confiança fornecidos pelos classificadores.
- Os níveis de confiança são somados para cada classe e a classe cuja soma resultante for a mais elevada, é declarada vencedora.

$$\hat{\omega} = \arg \max_{\omega \in \Omega} \sum_{i=1}^k P_{D_i}(\omega | x)$$

Soma Ponderada

- Ao invés de utilizar somente a regra da soma, é possível adicionar pesos às saídas dos classificadores, dando mais pesos aos classificadores mais competentes.
- Neste caso, o nível de confiança de cada classificador é multiplicado por k pesos w_i , e os pesos são específicos para cada classificador.

$$\hat{\omega} = \arg \max_{\omega \in \Omega} \sum_{i=1}^k w_i P_{D_i}(\omega | x)$$

Produto

- A regra do produto é baseada na multiplicação dos níveis de confiança fornecidos pelos classificadores.
- Os níveis de confiança são multiplicados, para cada classe, e a classe cujo produto resultante for o mais elevado, é declarada vencedora.

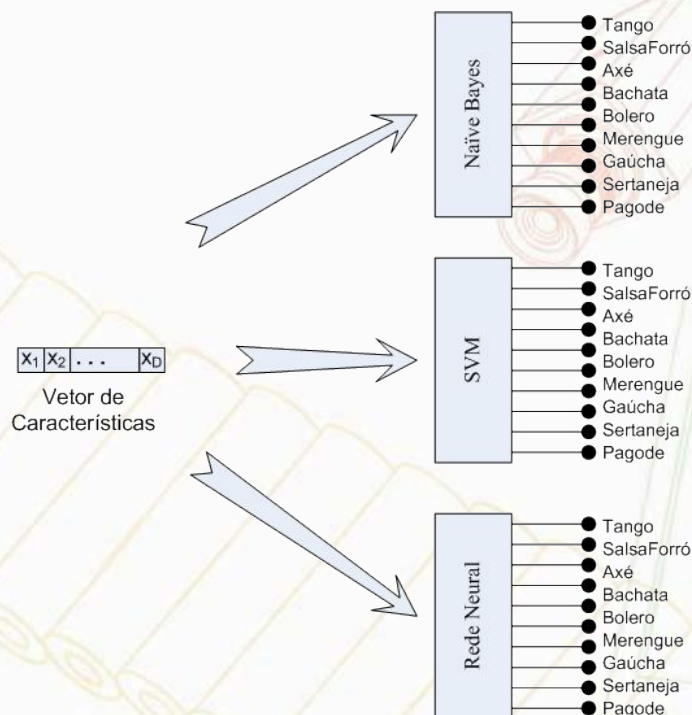
$$\hat{\omega} = \arg \max_{\omega \in \Omega} \prod_{i=1}^k P_{D_i}(\omega | x)$$

Produto Ponderado

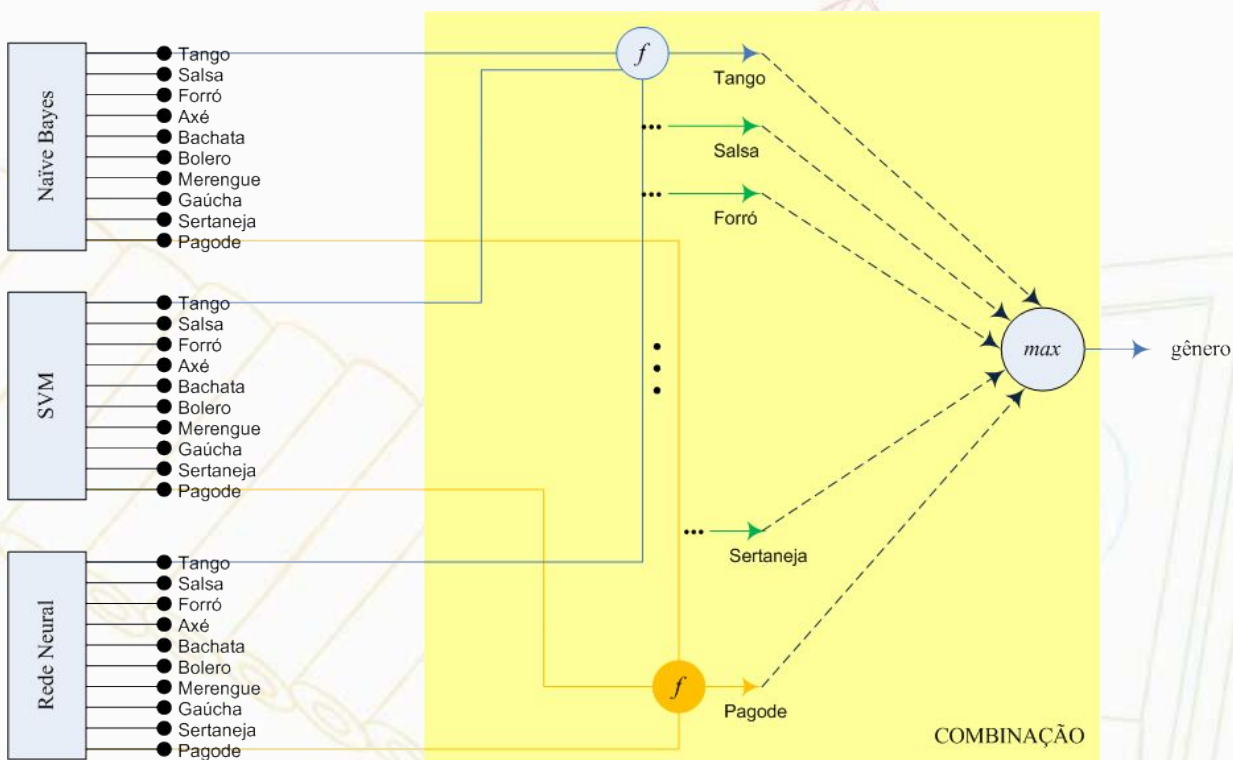
- Segue a mesma ideia da soma ponderada.
- Neste caso, o nível de confiança de cada classificador é elevado à k pesos w_i , e os pesos são específicos para cada classificador.

$$\hat{\omega} = \arg \max_{\omega \in \Omega} \prod_{i=1}^k [P_{D_i}(\omega | x)]^{w_i}$$

Exemplo



Exemplo



Exemplo

