



Machine Learning Boosting

Prof. Hugo de Paula

Boosting

Método para melhorar a precisão de qualquer algoritmo de aprendizado.

Funciona criando uma série de *datasets* tal que mesmo um desempenho modesto sobre essa base de dados pode ser utilizada para construir um previsor de alta precisão.

Adaboost (Adaptive Boosting)

Adaboost é considerado um dos *Top 10* do *Machine Learning*

Seja o conjunto de dados:

$S = \{(\mathbf{x}_1, y_1); (\mathbf{x}_2, y_2); \dots; (\mathbf{x}_m, y_m)\}$, onde $\mathbf{x} \in X, y \in [-1, 1]$

E um modelo de aprendizado fraco **A**.

Adaboost (Adaptive Boosting)

Para $t = 1, 2, \dots, T$

(1) Construir Domínio \mathbf{D}_t sobre $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$

(2) Executar \mathbf{A} sobre \mathbf{D}_t produzindo $\mathbf{H}_t: X \rightarrow Y$

(3) Seleciona modelo com menor erro

$$\varepsilon_t = P_{x_i \sim \mathbf{D}_t}(\mathbf{h}_t(x_i) \neq y_i), \text{ erro de } \mathbf{h}_t \text{ sobre } \mathbf{D}_t.$$

(3) Saída: $\mathbf{H} = \text{sign}(\sum_{t=1} \alpha_t \mathbf{h}_t(x))$

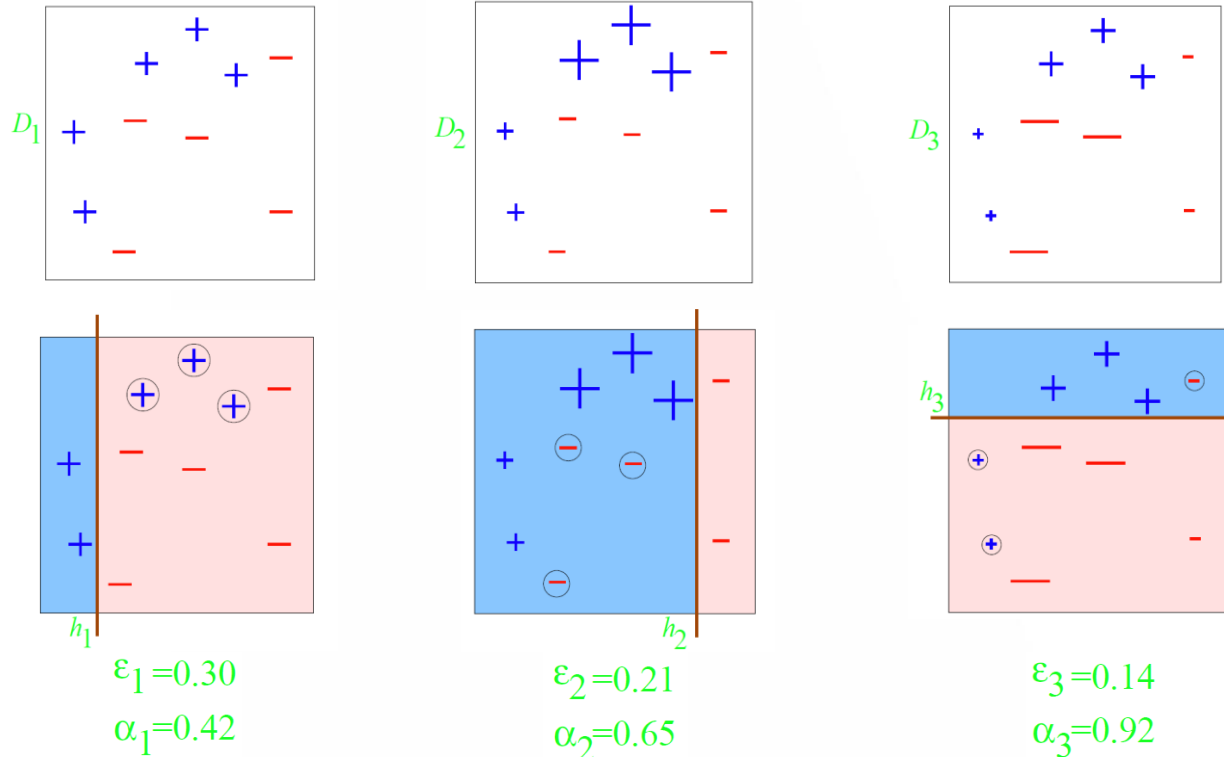
Características do Adaboost

Pode usar qualquer classificador fraco.

É rápido, pois faz apenas uma passada na base a cada iteração.

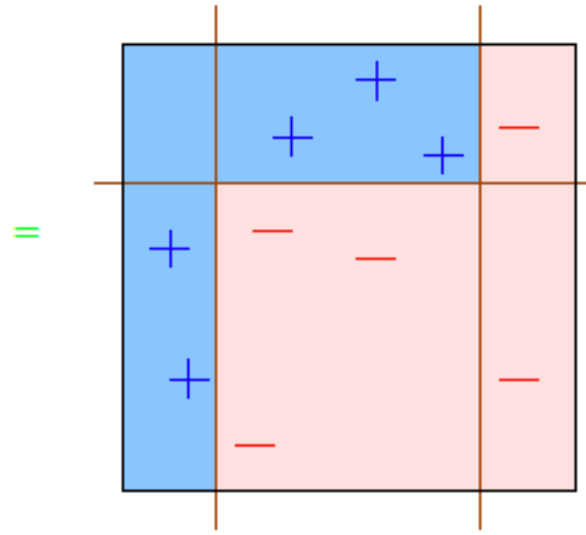
Mudança de mentalidade: O objetivo é encontrar um classificador que seja marginalmente melhor que adivinhação.

Adaboost: exemplo



H_{final} Adaboost: exemplo

$$= \text{sign} \left(0.42 \begin{array}{|c|} \hline \text{blue} \\ \hline \end{array} + 0.65 \begin{array}{|c|} \hline \text{blue} \\ \hline \end{array} + 0.92 \begin{array}{|c|} \hline \text{blue} \\ \hline \end{array} \right)$$



Exemplo: Base de dados Sonar (UCI, Scott E. Fahlman)

PROBLEMA: Baseado em dados de Sonar, prever se é rocha ou mina.

- 208 exemplos.
- 60 atributos numéricos.

Exemplo: Base de dados Sonar (UCI, Scott E. Fahlman)

RESULTADO: árvore de decisão

	Rocha	Mina	precisão
prev. Rocha	75	7	91.46%
prev. Mina	22	104	82.54%
revocação	77.32%	93.69%	

Acurácia: 86.06%

Precisão: 82.54% (classe: Mina), Revocação: 93.69%

Exemplo: Base de dados Sonar (UCI, Scott E. Fahlman)

RESULTADO: AdaBoost com Árvore de Decisão, 5 modelos

	Rocha	Mina	Precisão
prev. Rocha	92	4	95.83%
prev. Mina	5	107	95.54%
Revocação	94.85%	96.40%	

Acurácia: 95.67%

Precisão: 95.54% (classe: Mina), Revocação: 96.40%

Exemplo: Base de dados Sonar (UCI, Scott E. Fahlman)

RESULTADO: AdaBoost com Árvore de Decisão, 6 modelos

	Rocha	Mina	Precisão
prev. Rocha	97	0	100.00%
prev. Mina	0	111	100.00%
Revocação	100.00%	100.00%	

Acurácia: 100.00%

Precisão: 100.00% (classe: Mina), Revocação: 100.00%

Resumo sobre Boosting

Rápido, mas não tanto quando comparado com os outros métodos.

Simples e fácil de programar.

Pode combinar com qualquer algoritmo de treinamento.

Tende a evitar o *overfitting*

Aviso legal

O material presente nesta apresentação foi produzido a partir de informações próprias e coletadas de documentos obtidos publicamente a partir da Internet. Este material contém ilustrações adquiridas de bancos de imagens de origem privada ou pública, não possuindo a intenção de violar qualquer direito pertencente à terceiros e sendo voltado para fins acadêmicos ou meramente ilustrativos. Portanto, os textos, fotografias, imagens, logomarcas e sons presentes nesta apresentação se encontram protegidos por direitos autorais ou outros direitos de propriedade intelectual.

Ao usar este material, o usuário deverá respeitar todos os direitos de propriedade intelectual e industrial, os decorrentes da proteção de marcas registradas da mesma, bem como todos os direitos referentes a terceiros que por ventura estejam, ou estiveram, de alguma forma disponíveis nos slides. O simples acesso a este conteúdo não confere ao usuário qualquer direito de uso dos nomes, títulos, palavras, frases, marcas, dentre outras, que nele estejam, ou estiveram, disponíveis.

É vedada sua utilização para finalidades comerciais, publicitárias ou qualquer outra que contrarie a realidade para a qual foi concebido. Sendo que é proibida sua reprodução, distribuição, transmissão, exibição, publicação ou divulgação, total ou parcial, dos textos, figuras, gráficos e demais conteúdos descritos anteriormente, que compõem o presente material, sem prévia e expressa autorização de seu titular, sendo permitida somente a impressão de cópias para uso acadêmico e arquivo pessoal, sem que sejam separadas as partes, permitindo dar o fiel e real entendimento de seu conteúdo e objetivo. Em hipótese alguma o usuário adquirirá quaisquer direitos sobre os mesmos.

O usuário assume toda e qualquer responsabilidade, de caráter civil e/ou criminal, pela utilização indevida das informações, textos, gráficos, marcas, enfim, todo e qualquer direito de propriedade intelectual ou industrial deste material.



PUC Minas
Virtual

© PUC Minas • Todos os direitos reservados, de acordo com o art. 184 do Código Penal e com a lei 9.610 de 19 de fevereiro de 1998.
Proibidas a reprodução, a distribuição, a difusão, a execução pública, a locação e quaisquer outras modalidades de utilização sem a devida autorização da Pontifícia Universidade Católica de Minas Gerais.