Optimal Minimal Margin Maximization with Boosting

Allan Grønlund  Kasper Green Larsen  Alexander Mathiasen (me)
What is boosting?
What is boosting?
What is boosting?
What is boosting?
What is boosting?
What is boosting?
What is boosting?
What is boosting?
What is boosting?
What is boosting?
What is boosting?
What is boosting?
What is boosting?
No Overfitting?
No Overfitting?

Boosting

[Graph showing a curve that decreases over time with labels on the x-axis from 0 to 1000 and on the y-axis from 0 to 20.]

Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998
No Overfitting?

Boosting

# Hypotheses

Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998
No Overfitting?

Boosting

Classification Error vs. # Hypotheses

Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998
No Overfitting?

Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998

Boosting

Classification Error

Training Error

# Hypotheses
No Overfitting?

Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998

Diagram showing the relationship between the number of hypotheses and classification error, test error, and training error in the context of boosting.
No Overfitting?

Perfectly classify training data

Boosting

Classification Error vs # Hypotheses

Test Error = Training Error

Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998
No Overfitting?

Perfectly classify training data

Test error still improves!

Classification Error

Training Error

Test Error

# Hypotheses

Boosting

Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998
No Overfitting?

How do we explain this?

Boosting

Test error still improves!

Perfectly classify training data

Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998
An explanation by the minimal margin
An explanation by the minimal margin
An explanation by the minimal margin
An explanation by the minimal margin

* Not technically correct, see paper for definition.
An explanation by the minimal margin

\[
\Pr_{(x,y) \sim D}[f(x) \neq y] = O \left( \sqrt{\frac{\ln |H| \ln m}{\theta^2 m}} \right)
\]
An explanation by the minimal margin

Generalization error

$$\Pr_{(x,y) \sim D}[f(x) \neq y] = O \left( \sqrt{\frac{\ln |H| \ln m}{\theta^2 m}} \right)$$
An explanation by the minimal margin

Breiman 1998

\[ \Pr_{(x,y) \sim D} [f(x) \neq y] = O \left( \sqrt{\frac{\ln |H| \ln m}{\theta^2 m}} \right) \]

Generalization error
An explanation by the minimal margin

\[ \Pr_{(x,y) \sim D}[f(x) \neq y] = O \left( \sqrt{\frac{\ln |H| \ln m}{\theta^2 m}} \right) \]

Breiman 1998
An explanation by the minimal margin

Breiman 1998

\[
\Pr_{x,y \sim D}[f(x) \neq y] = O \left( \sqrt{\frac{\ln |H| \ln m}{\theta^2 m}} \right)
\]

Generalization error

Number of hypotheses

Number of points

Minimal margin
AdaBoostV: Finds a classifier with a provable bound on the trade-off between the number of hypotheses and the minimal margin. [Rätsch and Warmuth 2005]

\[
\Pr_{(x,y) \sim D}[f(x) \neq y] = O \left( \sqrt{\frac{\ln |H| \ln m}{\theta^2 m}} \right)
\]
An explanation by the minimal margin

Generalization error

\[ \Pr_{(x,y) \sim D}[f(x) \neq y] = O\left(\sqrt{\frac{\ln |H| \ln m}{\theta^2 m}}\right) \]

AdaBoostV: Finds a classifier with a provable bound on the trade-off between the number of hypotheses and the minimal margin. [Rätsch and Warmuth 2005]

JMLR: Conjecture: there is a lower bound which matches AdaBoostV. [Nie et al. 2013]
An explanation by the minimal margin

Breiman 1998

Generalization error

Number of hypotheses

Number of points

Minimal margin

$\Pr_{(x,y) \sim D} [f(x) \neq y] = O \left( \sqrt{\frac{\ln |H| \ln m}{\theta^2 m}} \right)$

AdaBoostV: Finds a classifier with a provable bound on the trade-off between the number of hypotheses and the minimal margin. [Rätsch and Warmuth 2005]

JMLR: Conjecture: there is a lower bound which matches AdaBoostV. [Nie et al. 2013]

FALSE.
An explanation by the minimal margin

AdaBoostV: Finds a classifier with a provable bound on the trade-off between the number of hypotheses and the minimal margin. [Rätsch and Warmuth 2005]

JMLR: Conjecture: there is a lower bound which matches AdaBoostV. [Nie et al. 2013] FALSE.

SparsiBoost: Obtains a slightly better bound than AdaBoostV.

\[ \Pr_{(x,y) \sim D}[f(x) \neq y] = O \left( \sqrt{\frac{\ln |H| \ln m}{\theta^2 m}} \right) \]
An explanation by the minimal margin

AdaBoostV: Finds a classifier with a provable bound on the trade-off between the number of hypotheses and the minimal margin. [Rätsch and Warmuth 2005]

JMLR: Conjecture: there is a lower bound which matches AdaBoostV. [Nie et al. 2013] FALSE.

SparsiBoost: Obtains a slightly better bound than AdaBoostV. Proved matching lower bound, so SparsiBoost is optimal.

\[ \Pr_{(x,y) \sim D} [f(x) \neq y] = O \left( \sqrt{\ln |H| \ln m \over \theta^2 m} \right) \]
Experiments