Optimal Minimal Margin Maximization with Boosting

Allan Grønlund

Kasper Green Larsen

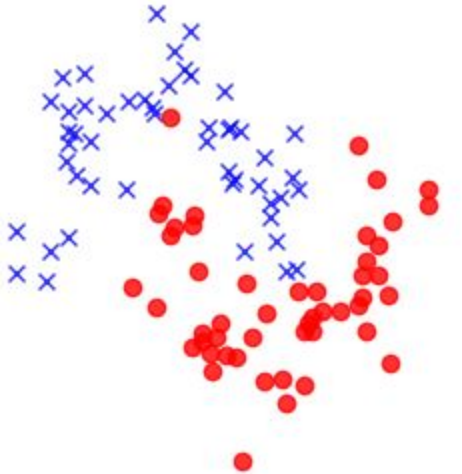
Alexander Mathiasen (me)

AU AARHUS UNIVERSITY DENMARK

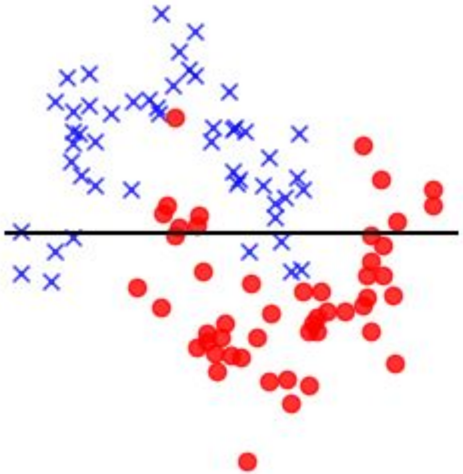
What is boosting?

What is boosting?

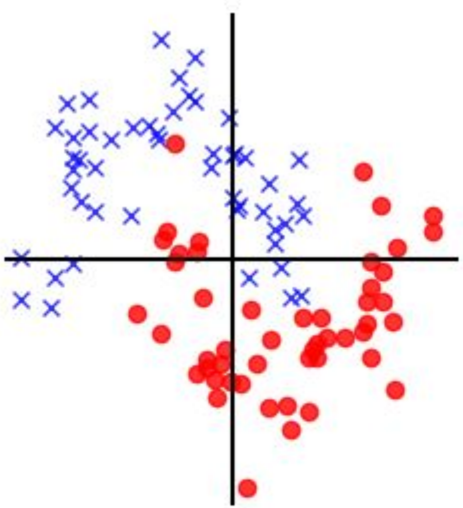
What is boosting?



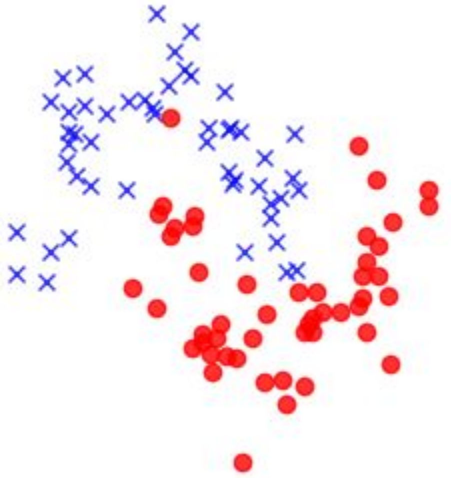
What is boosting?



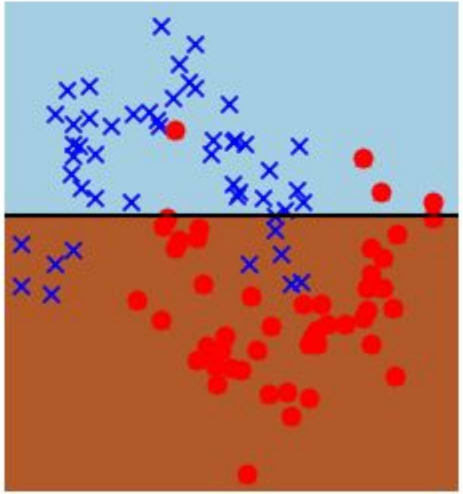
What is boosting?

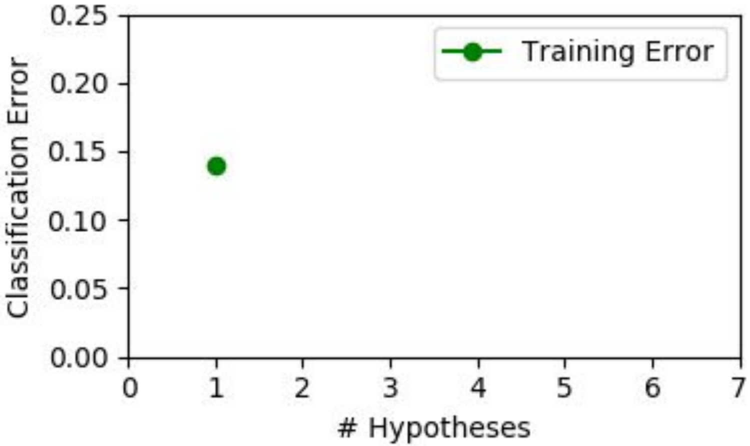


What is boosting?

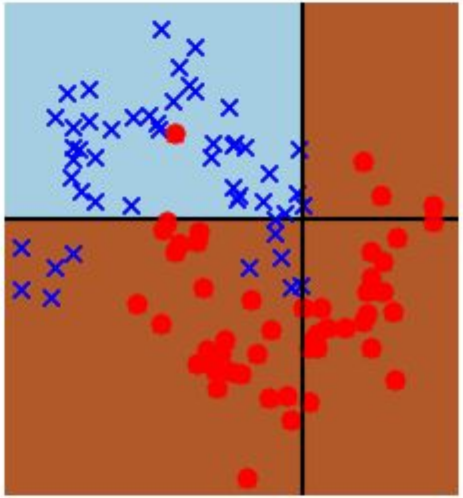


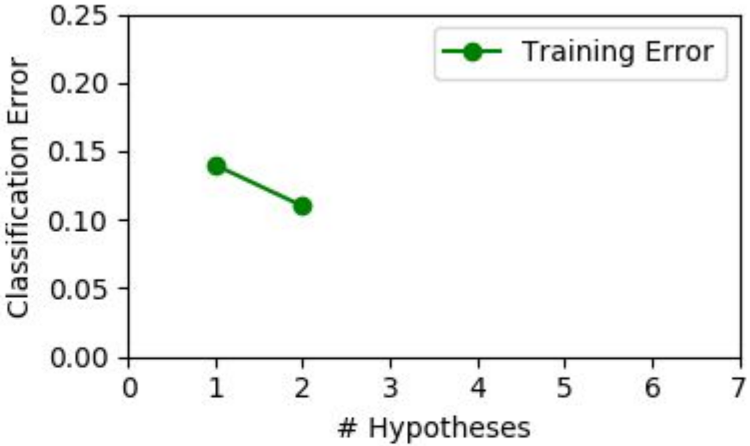
What is boosting?



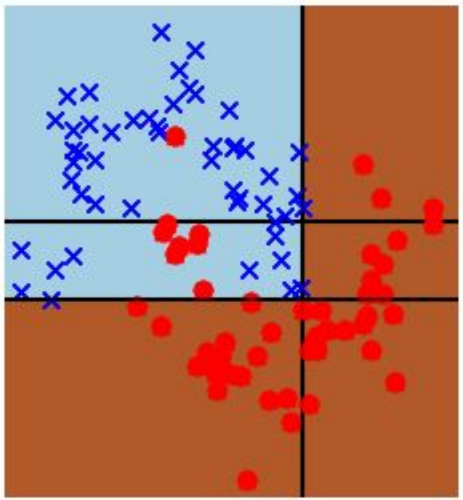


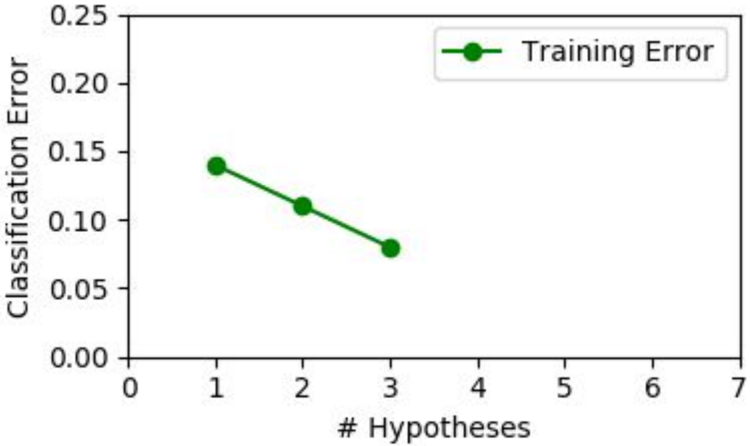
What is boosting?



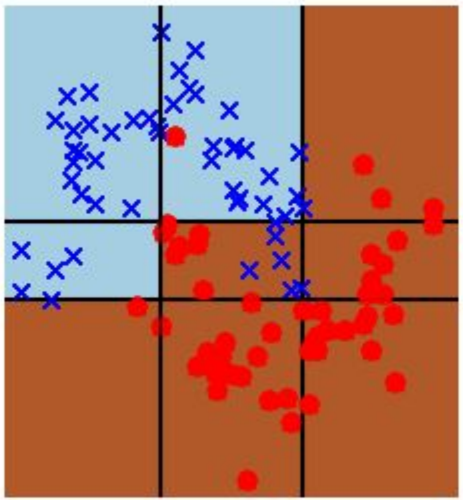


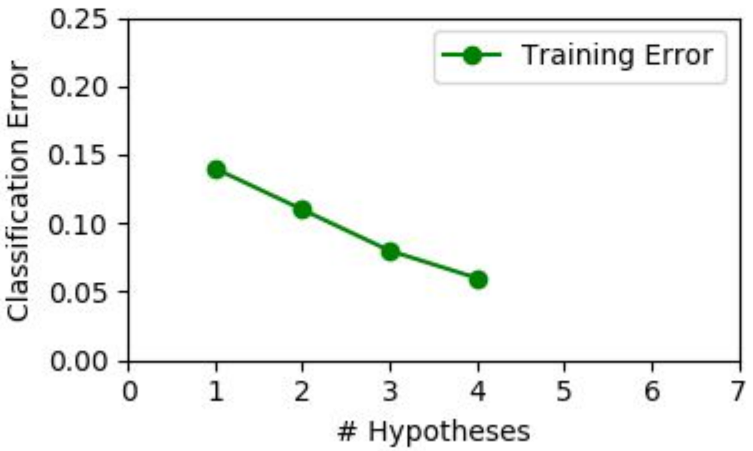
What is boosting?



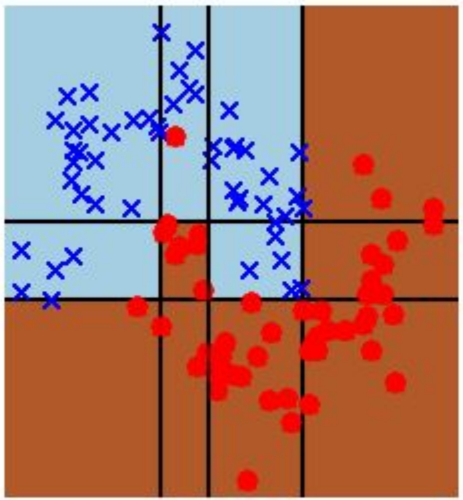


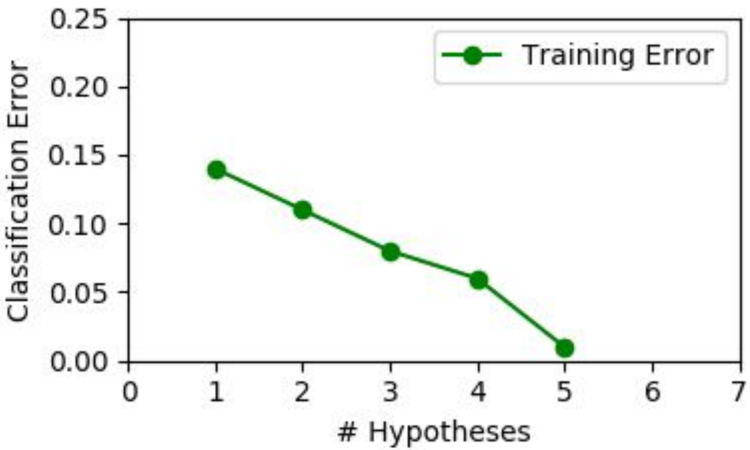
What is boosting?



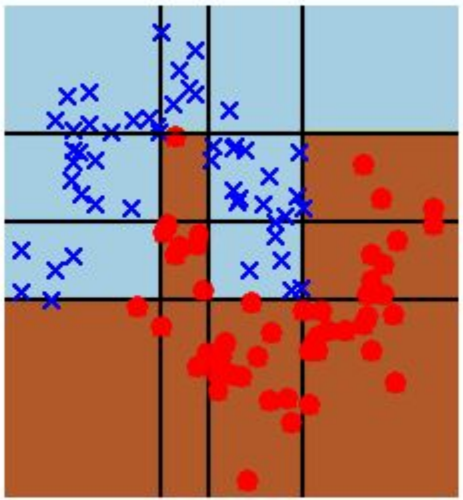


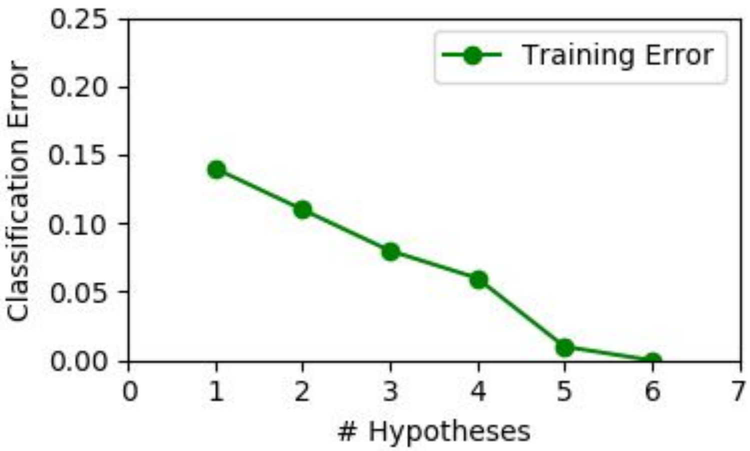
What is boosting?



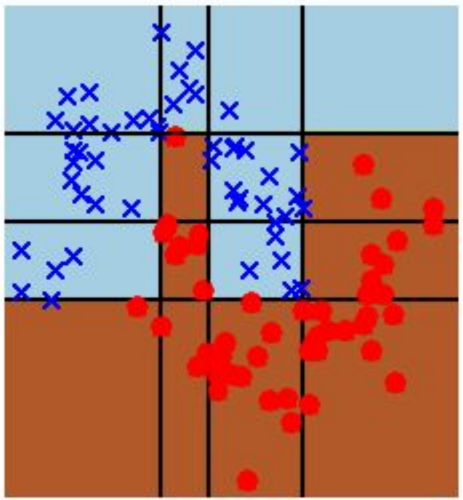


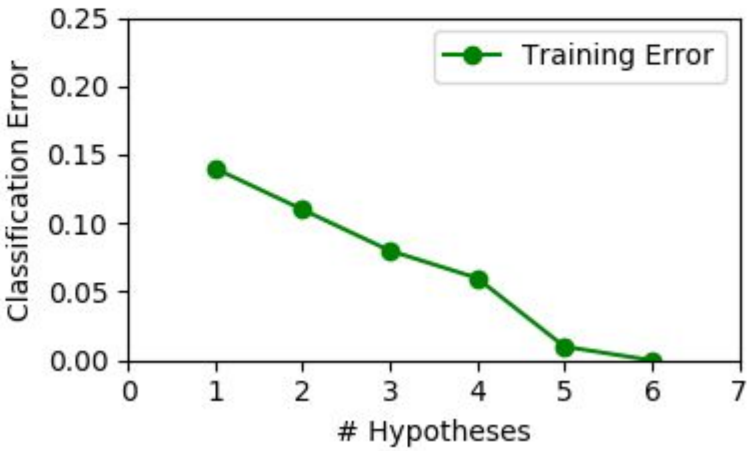
What is boosting?





What is boosting?

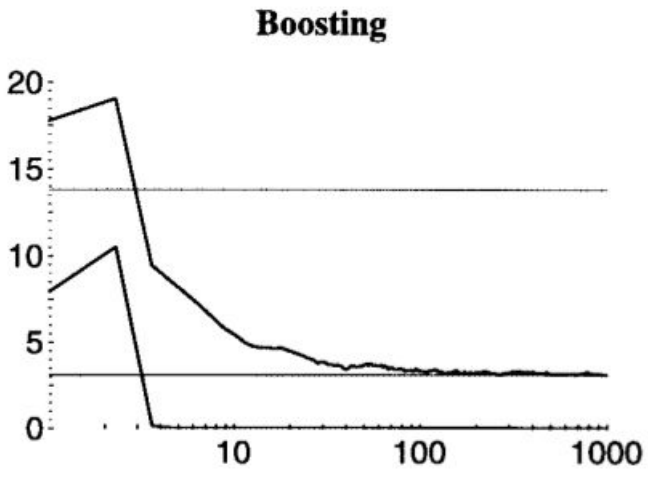




No Overfitting?

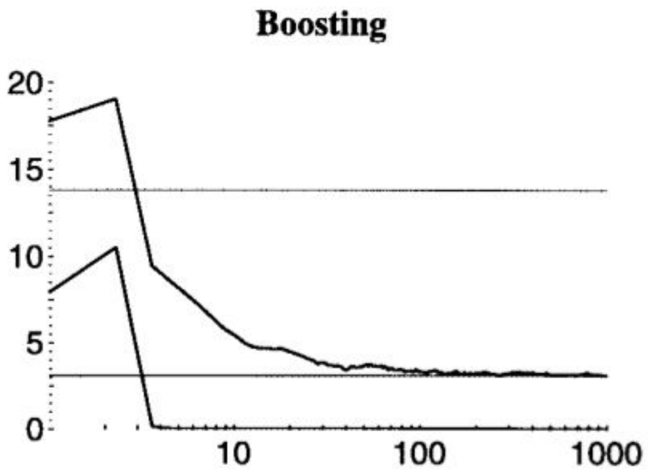
Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998

No Overfitting?



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

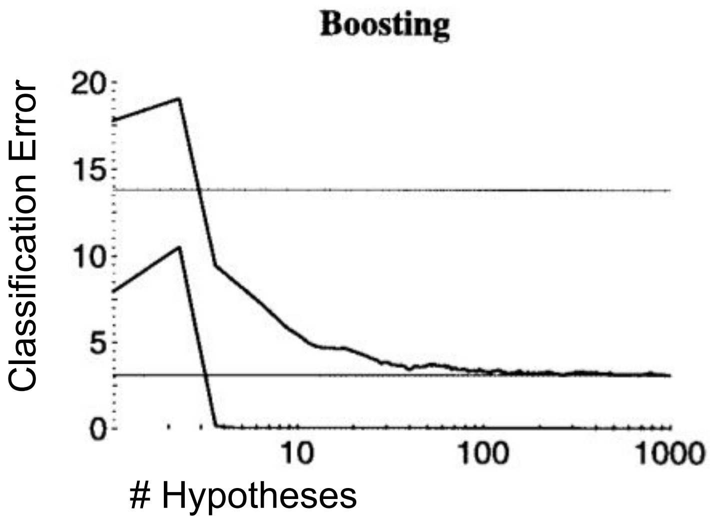
No Overfitting?



Hypotheses

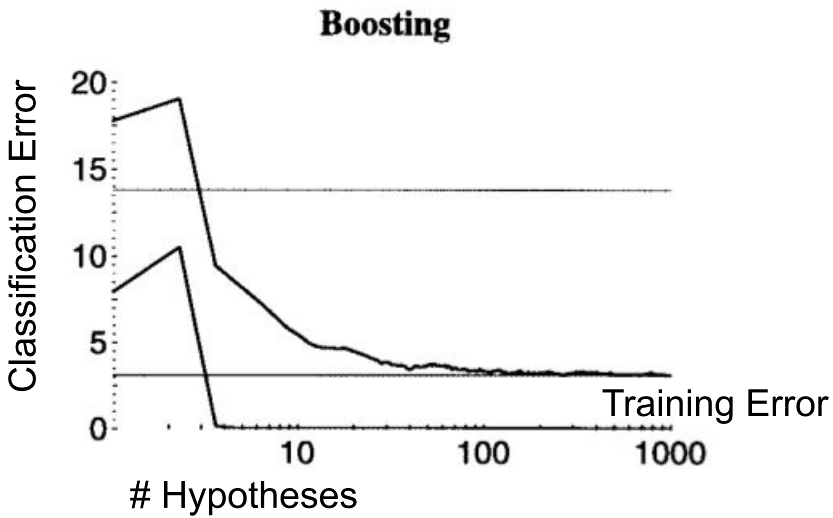
Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998

No Overfitting?



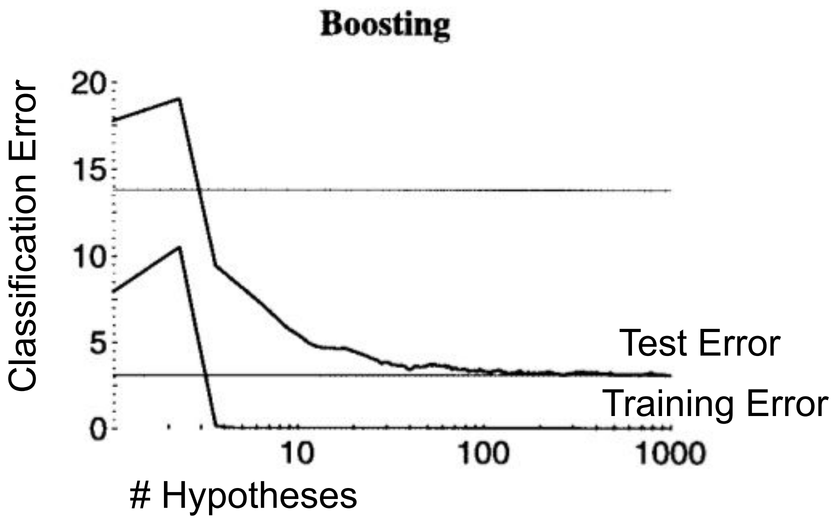
Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998

No Overfitting?



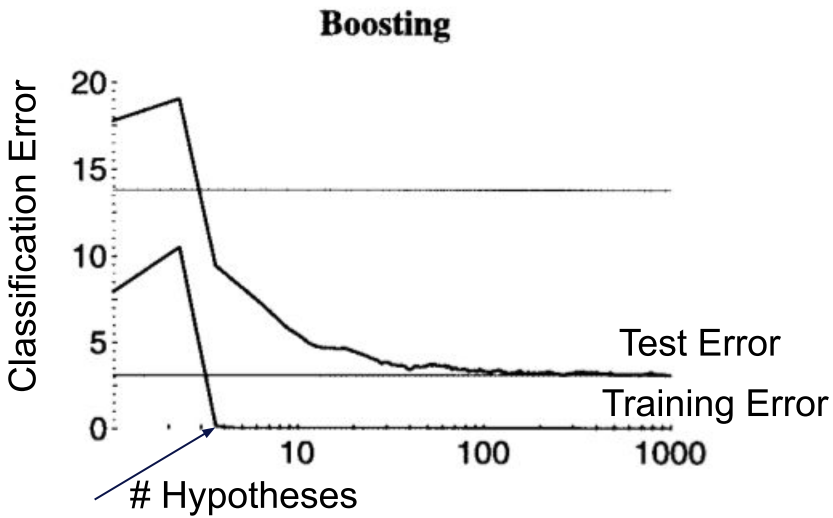
Schapire, Robert E.; Freund, Yoav; Bartlett, Peter; Lee, Wee Sun 1998

No Overfitting?



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

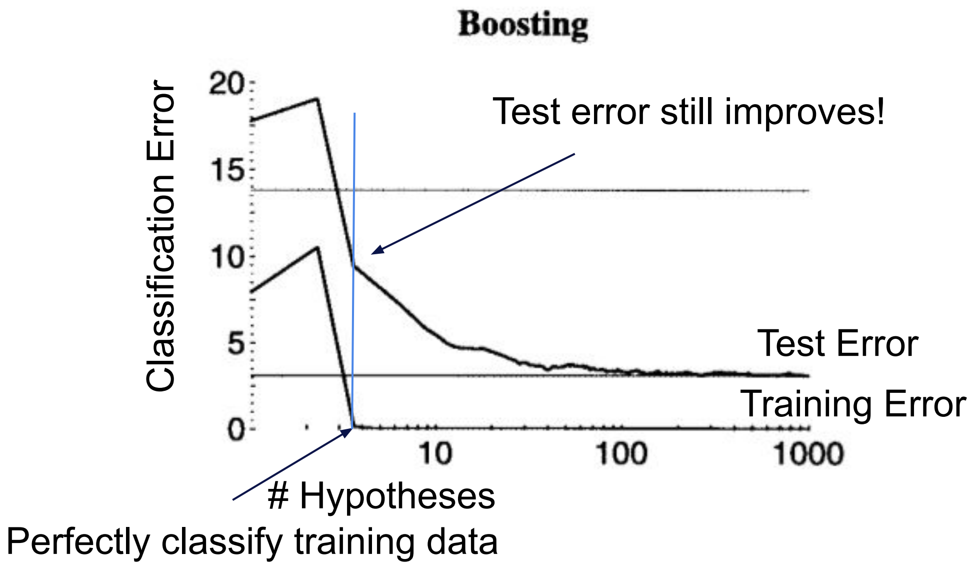
No Overfitting?



Perfectly classify training data

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

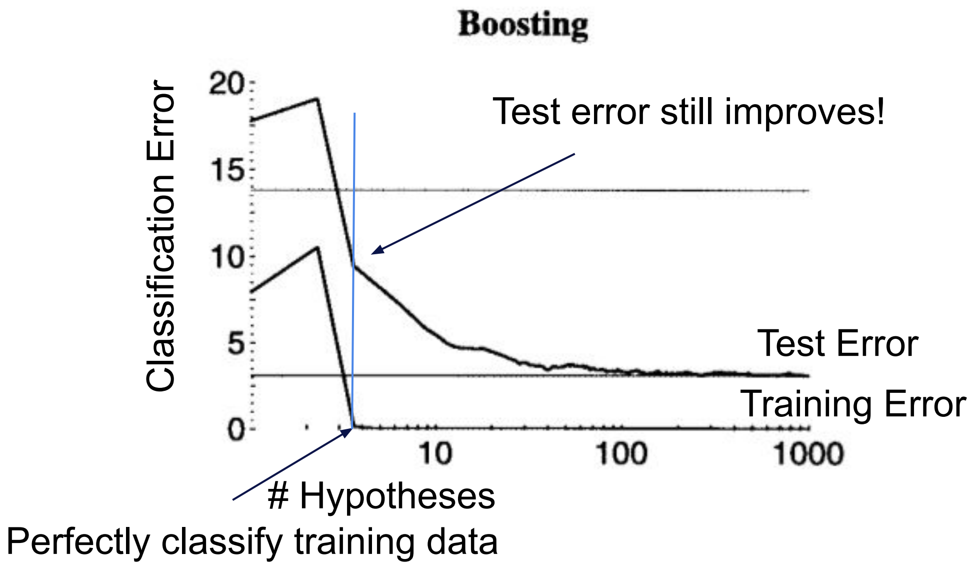
No Overfitting?



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

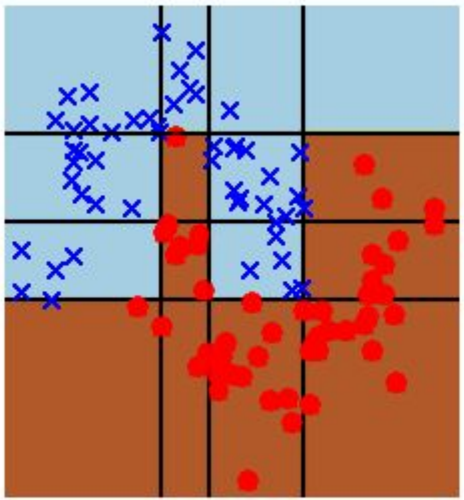
No Overfitting?

How do we explain this?

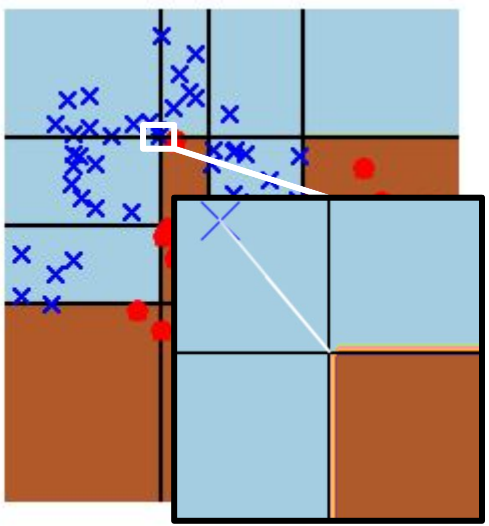


An explanation by the minimal margin

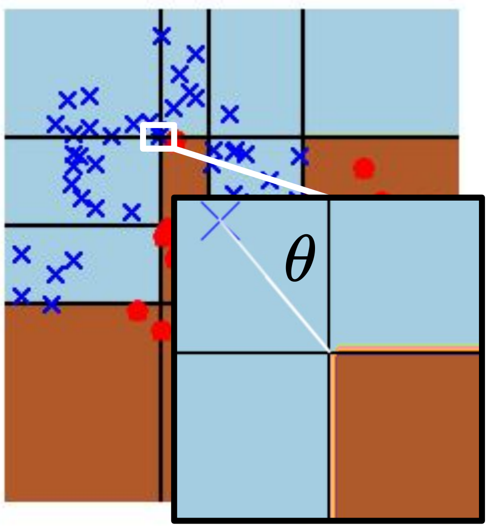
An explanation by the minimal margin



An explanation by the minimal margin



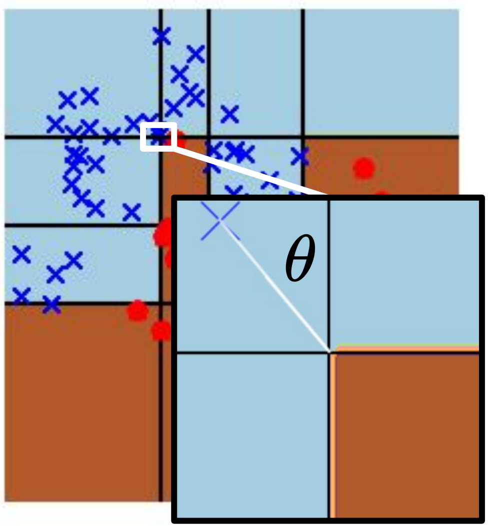
An explanation by the minimal margin



* *Not technically correct, see paper for definition.*

Breiman 1998

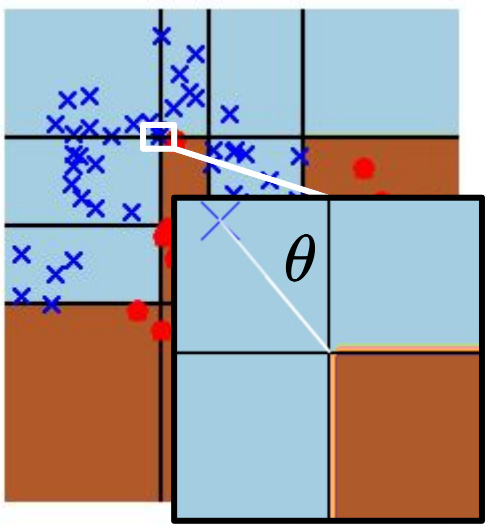
An explanation by the minimal margin



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

Breiman 1998

An explanation by the minimal margin

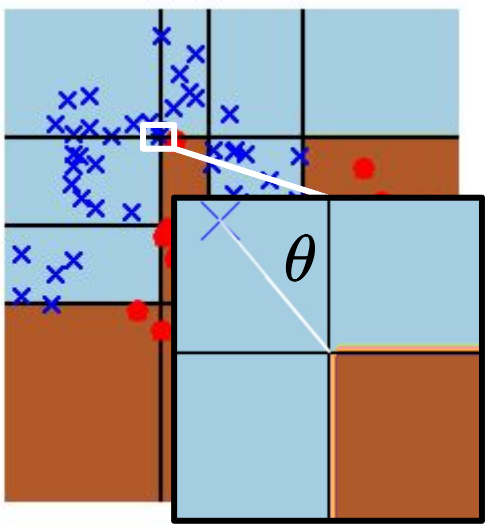


Generalization error

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

Breiman 1998

An explanation by the minimal margin



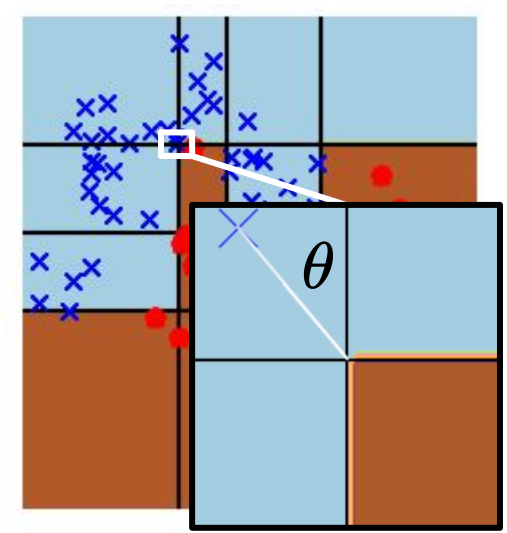
Generalization error

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

Minimal margin

Breiman 1998

An explanation by the minimal margin



Generalization error

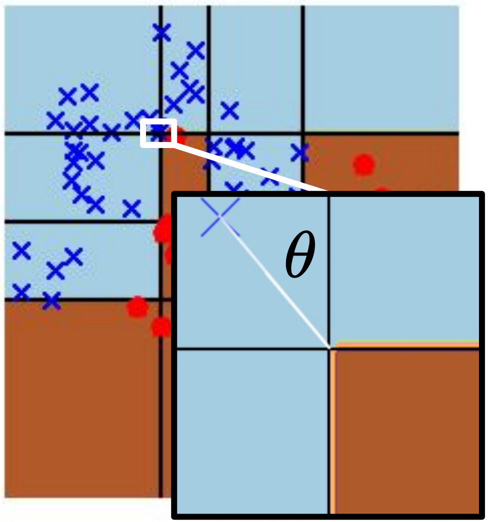
Number of hypotheses

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

Minimal margin

Breiman 1998

An explanation by the minimal margin



Generalization error

Number of hypotheses

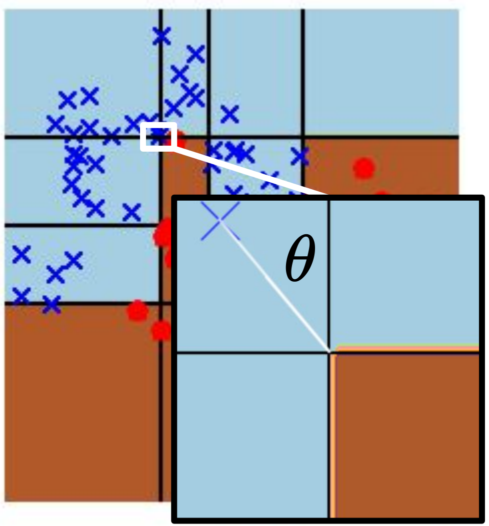
Number of points

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

Minimal margin

Breiman 1998

An explanation by the minimal margin



Generalization error

Number of hypotheses

Number of points

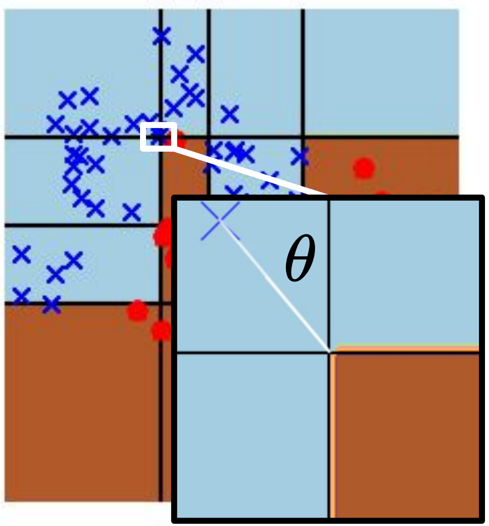
$$\Pr_{(x,y)\sim D}[f(x) \neq y] = O\left(\sqrt{\frac{\ln|H|\ln m}{\theta^2 m}}\right)$$

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]

Breiman 1998

An explanation by the minimal margin



Generalization error

Number of hypotheses

Number of points

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

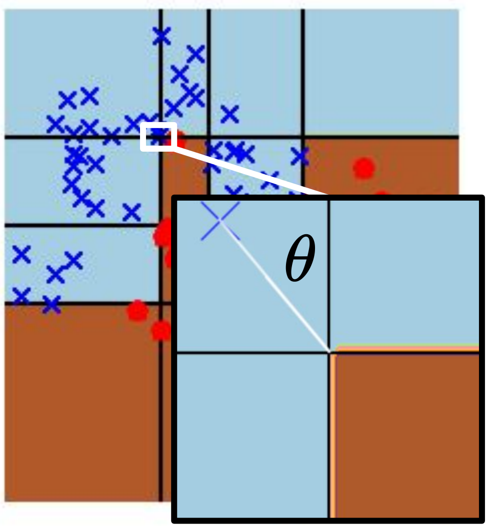
Minimal margin

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]

Breiman 1998

An explanation by the minimal margin



Generalization error

Number of hypotheses

Number of points

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

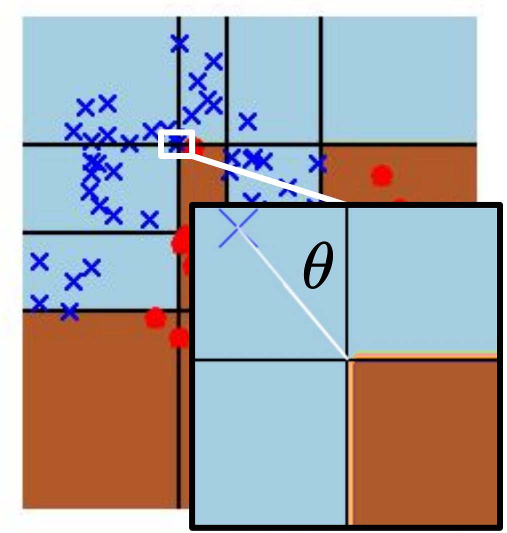
Minimal margin

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.

Breiman 1998

An explanation by the minimal margin



Number of hypotheses

Number of points

Generalization error

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

Minimal margin

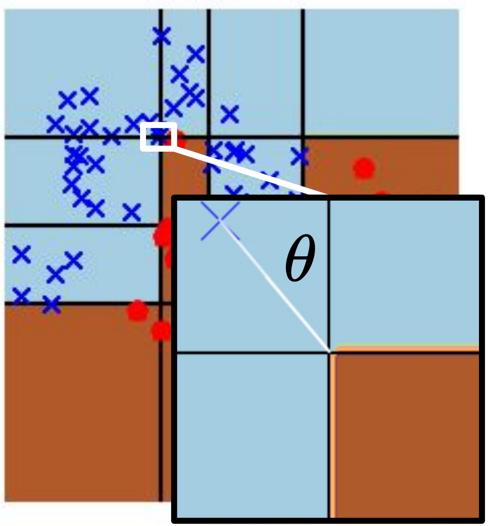
AdaBoostV: Finds a classifier with a provable bound on the trade-off between the number of hypotheses and the minimal margin. [Rätsch and Warmuth 2005]

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

SparsiBoost: Obtains a slightly better bound than AdaBoostV.

Breiman 1998

An explanation by the minimal margin



Generalization error

Number of hypotheses

Number of points

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

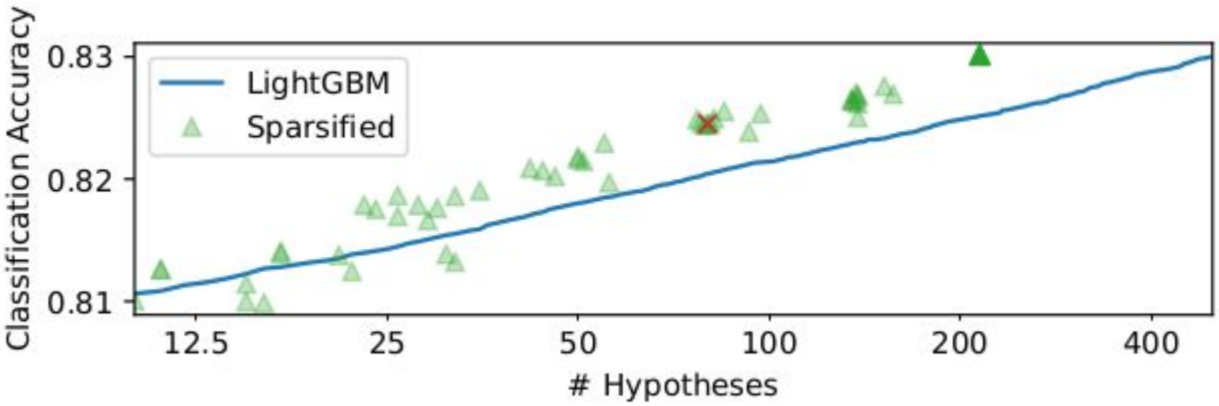
Minimal margin

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.

Experiments



AU

AARHUS
UNIVERSITY
DENMARK