AdaBoost

Boosting algorithm

1. Initialize $f(x) = 0, r = y$

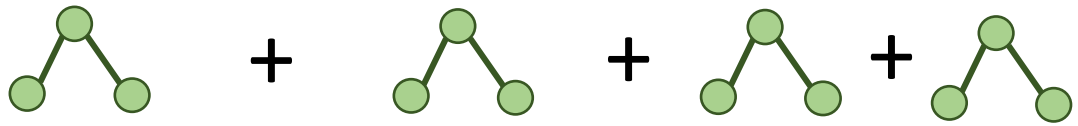
2. For $b = 1,2, \ldots, B$, repeat

   a) Fit a tree $f_b(x)$ to the training data (X, r)

   b) $f(x) \leftarrow f(x) + \lambda f_b(x)$

   c) $r \leftarrow r - \lambda f_b(x)$

3. output $\sum_{b=1}^{B} \lambda f_b(x)$



Popular Boosting Algorithms

- AdaBoost
- Gradient Boost

AdaBoost

- Classification
- Uses weights on data samples
- Each stump fits to y
- Predicts discrete values -1 or 1
- Uses exponential weight update

AdaBoost

1. Initialize sample weights $w_i = 1/N$

$$F(x) = \text{sign}\left[\sum_{b=1}^{B} \lambda_b f_b(x)\right]$$

2. For $b = 1,2,\ldots,B$, repeat

   a) Fit a tree $f_b(x)$ to the training data (X, y) with sample weights w_i

   b) Calculate error $\varepsilon_b = \frac{\sum_{i=1}^{N} w_i I(y_i \neq f_b(x_i))}{\sum_{i=1}^{N} w_i}$

   c) Calculate model coefficient $\lambda_b = \frac{1}{2}\log\left(\frac{1-\varepsilon_b}{\varepsilon_b}\right)$

   d) Update sample weights $w_i \leftarrow w_i \exp\left(\lambda_b \cdot I(y_i \neq f_b)(x_i)\right)$

3. output $F(x) = \text{sign}\left[\sum_{b=1}^{B} \lambda_b f_b(x)\right]$

AdaBoost

1. Initialize sample weights $w_i = 1/N$

	Age	Sex	ChestPain	Chol	AHD	weight
0	63	1	typical	233	No	0.1
1	67	1	asymptomatic	286	Yes	0.1
2	67	1	asymptomatic	229	Yes	0.1
3	37	1	nonanginal	250	No	0.1
4	41	0	nontypical	204	No	0.1
5	56	1	nontypical	236	No	0.1
6	62	0	asymptomatic	268	Yes	0.1
7	57	0	asymptomatic	354	No	0.1
8	63	1	asymptomatic	254	Yes	0.1
9	53	1	asymptomatic	203	Yes	0.1

AdaBoost

2. For $b = 1,2, \ldots, B$, repeat

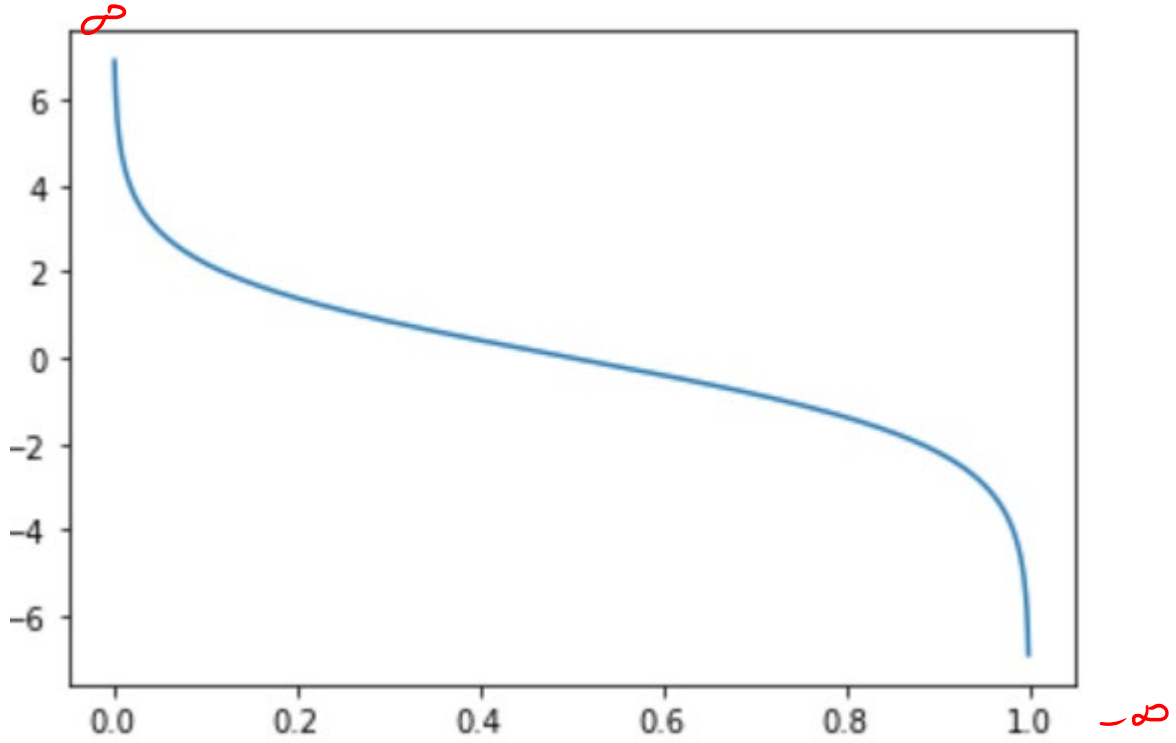
   a) Fit a tree $f_b(x)$ to the training data (X, y) with sample weights w_i

	Age	Sex	ChestPain	Chol	AHD	weight	Yp
0	63	1	typical	233	No	0.1	Yes
1	67	1	asymptomatic	286	Yes	0.1	Yes
2	67	1	asymptomatic	229	Yes	0.1	Yes
3	37	1	nonanginal	250	No	0.1	No
4	41	0	nontypical	204	No	0.1	No
5	56	1	nontypical	236	No	0.1	No
6	62	0	asymptomatic	268	Yes	0.1	Yes
7	57	0	asymptomatic	354	No	0.1	No
8	63	1	asymptomatic	254	Yes	0.1	Yes
9	53	1	asymptomatic	203	Yes	0.1	No

AdaBoost

2-b) Calculate Error $\varepsilon_b = \frac{\Sigma_{i=1}^{N} w_i I(y_i \neq f_b(x_i))}{\Sigma_{i=1}^{N} w_i}$ = 0.2

2-c) Calculate model coefficient $\lambda_b = \log(\frac{1-\varepsilon_b}{\varepsilon_b})$ = 1.386



AdaBoost

2-d) Update sample weights $w_i \leftarrow w_i \exp\left(\lambda_b \cdot I\left(y_i \neq f_b(x)\right)\right)$

	Age	Sex	ChestPain	Chol	AHD	weight	Yp
0	63	1	typical	233	No	4.0	Yes
1	67	1	asymptomatic	286	Yes	1.0	Yes
2	67	1	asymptomatic	229	Yes	1.0	Yes
3	37	1	nonanginal	250	No	1.0	No
4	41	0	nontypical	204	No	1.0	No
5	56	1	nontypical	236	No	1.0	No
6	62	0	asymptomatic	268	Yes	1.0	Yes
7	57	0	asymptomatic	354	No	1.0	No
8	63	1	asymptomatic	254	Yes	1.0	Yes
9	53	1	asymptomatic	203	Yes	4.0	No

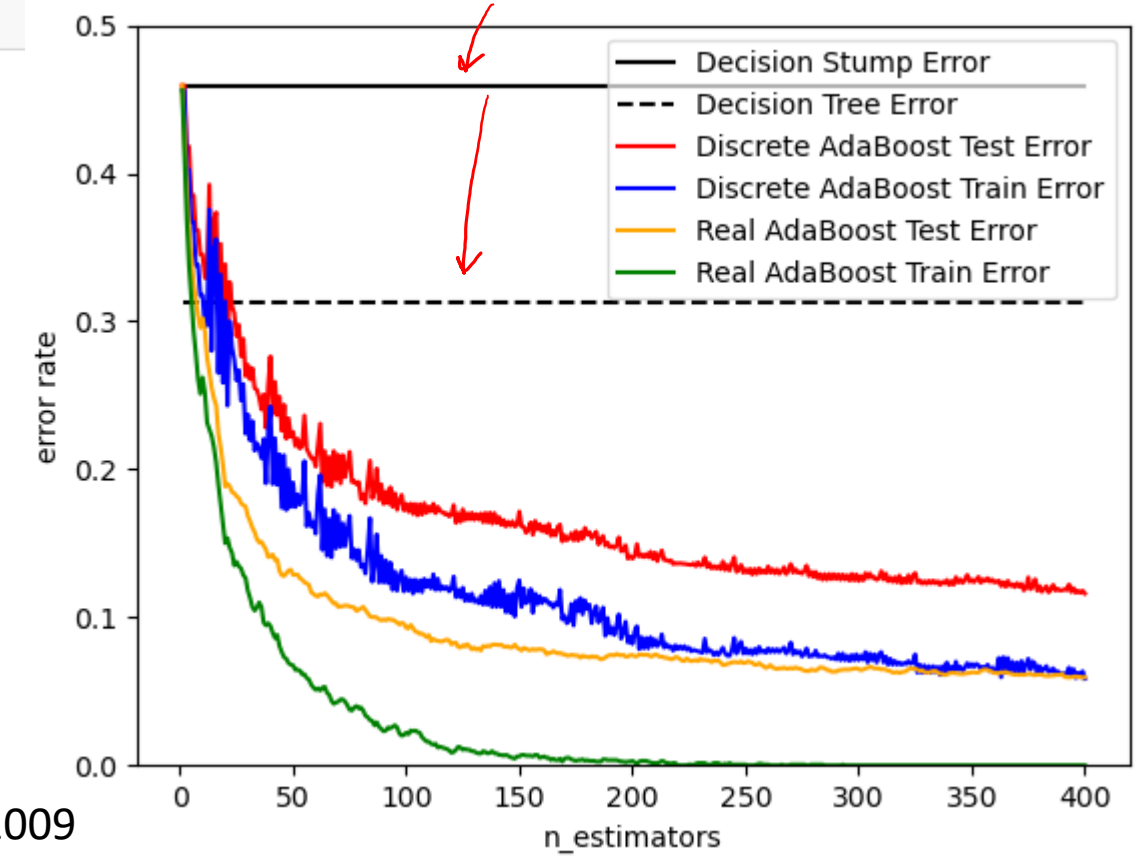
AdaBoost

	Age	Sex	ChestPain	Chol	AHD	weight	Yp
0	63	1	typical	233	No	0.2500	Yes
1	67	1	asymptomatic	286	Yes	0.0625	Yes
2	67	1	asymptomatic	229	Yes	0.0625	Yes
3	37	1	nonanginal	250	No	0.0625	No
4	41	0	nontypical	204	No	0.0625	No
5	56	1	nontypical	236	No	0.0625	No
6	62	0	asymptomatic	268	Yes	0.0625	Yes
7	57	0	asymptomatic	354	No	0.0625	No
8	63	1	asymptomatic	254	Yes	0.0625	Yes
9	53	1	asymptomatic	203	Yes	0.2500	No

AdaBoost

sklearn.ensemble.AdaBoostClassifier

class `sklearn.ensemble.` **`AdaBoostClassifier`**(*base_estimator=None, *, n_estimators=50, learning_rate=1.0, algorithm='SAMME.R', random_state=None*)



Zhu, H. Zou, S. Rosset, T. Hastie, "Multi-class AdaBoost", 2009

https://scikit-learn.org/stable/auto_examples/ensemble/plot_adaboost_hastie_10_2.html#id4

AdaBoost

Can AdaBoost also do Regression?

sklearn.ensemble.AdaBoostRegressor

class sklearn.ensemble.**AdaBoostRegressor**(*base_estimator=None, *, n_estimators=50, learning_rate=1.0, loss='linear', random_state=None*) [source]

How good is AdaBoost?

- Data: ~5k+ samples, ~20 features