Boosting:

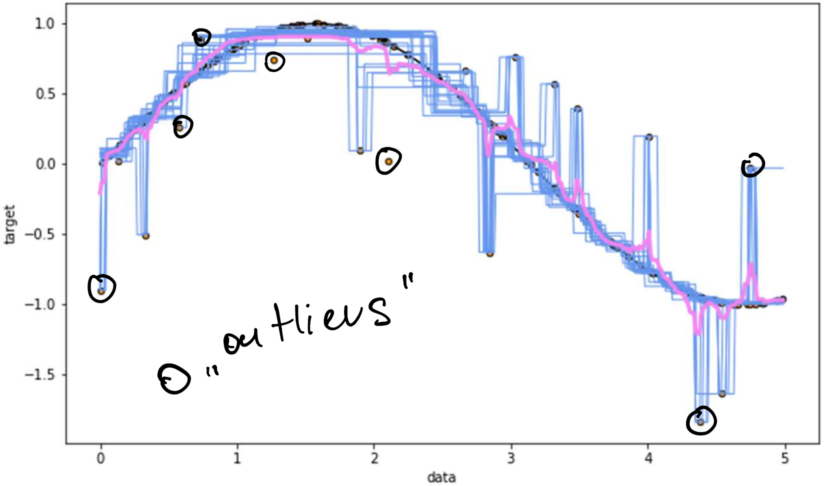
Greedy reducing errors

Bagging

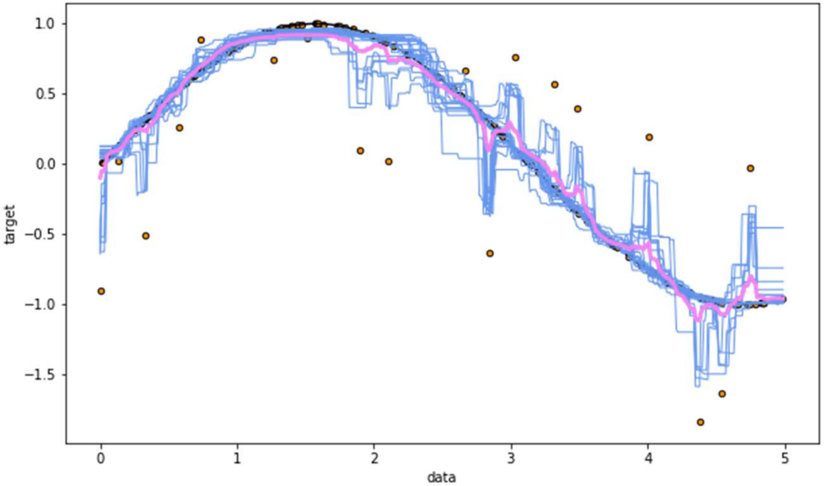
ensemble of N base models

- Bias $a_N(x)$ is the same $b_n(x)$
- Variance $a_N(x)$:
- $\frac{1}{N}$(Variance $b_n(x)$) $+ \sum_{n<m}$ cov $(b_n(x), b_m(x))$
- 1) cov$(b_n, b_m) \approx 0$, Variance ↓ N times
- 2) the much as this true, then better.

Trees



Several
Bagged Trees



Problems
(Trees)
≈ 500

more, more

1) b_n is biased $\Rightarrow \frac{1}{N}\sum_{n=1}^{N} b_n$ has the same bias

2) Learning large $\underline{\underline{N}}$

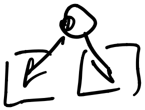
✓ Ok
- independ learning
$\Rightarrow$ $\underline{\underline{\parallel}}$

×
- independantley
large N $\Rightarrow$
- costs to store
- costs to apply
- lack of interpetation

Idea Boosting

(opposite for bagging)

Boost

- Very $\underbrace{\text{simple}}_{\text{big bias}\uparrow,\ \text{low variance}\downarrow}$ models 
-
- $\underbrace{\text{Sequentially}}$
-

each new model correct previos

Bagging

Very $\underbrace{\text{complex}}_{\text{big variance}\uparrow,\ \text{small bias}\downarrow}$ models

Independent, ||

ensemble

base algorithms

$$\underbrace{a_N(x)} = \sum_{n=1}^{N} b_n(x)$$

- How to learn the (b_1):

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L(y_i, b_1(x_i)) \to \min_{b_1(x)}$$

loss function

train dataset

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

- b_1):

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(y_i, b_1(x_i)\big) \to \min_{b_1(x)}$$

- b_1,

Next models?

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

- Learn N model (b_N): ($b_{N-1}, \dots, b_1$ fixed)

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L(y_i, \underbrace{a_{N-1}(x_i)}_{\text{fixed already}} + \underbrace{b_N(x_i)}_{\text{learn}}) \to \min_{b_N(x)}$$

1) LR - ∇_w
2) Tree?

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

- N (b_N) :

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(y_i, a_{N-1}(x_i) + b_N(x_i)\big) \to \min_{b_N(x)}$$

- b_N

$$a_N = a_{N-1} + b_N$$

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

- N (b_N) :

$$\frac{1}{\ell} \sum_{i=1}^{\ell} L\left(y_i, a_{N-1}(x_i) + b_N(x_i)\right) \to \min_{b_N(x)}$$

- Learning N model (b_N):

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(y_i, a_{N-1}(x_i) + b_N(x_i)\big) \to \min_{b_N(x)}$$

?• Don't know how to optimize over b_N

?• Introduce modifications, but suitable for any model

Outline

- Learn sequentially
- New models | Old models and "try" to correct
- We need new training routine

[Opposite to Bagging Idea]

Boosting for

Mean Squared Error

(MSE)

=> regression task

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

- New N model:

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L(y_i, \underbrace{a_{N-1}(x_i) + b_N(x_i)}) \to \min_{b_N(x)}$$

MSE

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

- Opt. N model :

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\Big(y_i, \underbrace{\overbrace{a_{N-1}(x_i)}^{\text{fixed}} + \overbrace{b_N(x_i)}^{\text{train}}}\Big) \to \min_{b_N(x)}$$

- MSE:

$$\underbrace{L(y,\hat{y}) = (\underset{\text{observed}}{y} - \underset{\text{prediction}}{\hat{y}})^2}$$

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

- N :

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\big(\underbrace{a_{N-1}(x_i) + b_N(x_i)}_{\text{output}} - \underbrace{y_i}_{\text{observe}}\big)^2 \to \min_{b_N(x)}$$

GN

matching residuals = (reduce errors sequentially)

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

0 if already no error

- N :

"

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\Big(b_N(x_i) - (\overbrace{y_i - a_{N-1}(x_i)}^{s_i})\Big)^2 \to \min_{b_N(x)}$$

$$= \big(y_i - (a_{N-1}(x) + b_N(x_i))\big)^2$$

$s_i = y_i - a_{N-1}(x_i)$ residuals

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

- N :

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\Big(b_N(x_i) - \underbrace{\big(y_i - a_{N-1}(x_i)\big)}_{s_i^{(N)}}\Big)^2 \to \min_{b_N(x)}$$

$$s_i^{(N)} = y_i - a_{N-1}(x_i) -$$

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left(b_N(x_i) - s_i^{(N)}\right)^2 \to \min_{b_N(x)}$$

- $s_i^{(N)} = y_i - a_{N-1}(x_i)$

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left(b_N(x_i) - s_i^{(N)}\right)^2 \to \min_{b_N(x)}$$

- $s_i^{(N)} = y_i - a_{N-1}(x_i)$
- b_N $s_i^{(N)}$,

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left(b_N(x_i) - s_i^{(N)}\right)^2 \to \min_{b_N(x)}$$

- $s_i^{(N)} = y_i - a_{N-1}(x_i) -$
- b_N $s_i^{(N)}$,

$$y_i = a_{N-1}(x_i) + s_i^{(N)} = a_{N-1}(x_i) + b_N(x_i)$$

Example

- $y_i = \underbrace{12}$
- $a_{N-1}(x_i) = \underbrace{10}$
- $s_i^{(N)} = ?$ 12 - 10 = 2
- $b_N(x_i) = ?$ + 2
- $a_N(x_i) = ?$ 12.

- $y_i = 12$
- $a_{N-1}(x_i) = 10$
- $s_i^{(N)} = 2$
- $b_N(x_i) = 2$
- $a_N(x_i) = 12$

First Iter.

$$\frac{1}{\ell}\sum_{i=1}^{\ell}(b_1(x_i) - y_i)^2 \to \min_{b_1(x)}$$ ✓

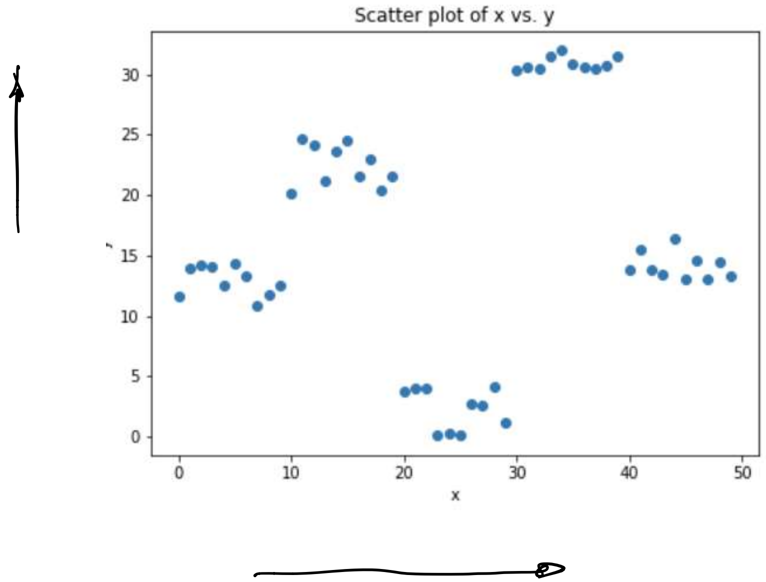
2 iteration

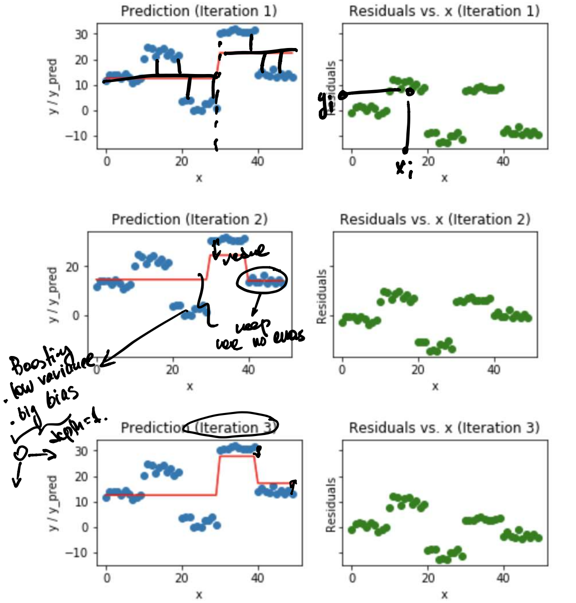
$$\frac{1}{\ell}\sum_{i=1}^{\ell}\Big(\underbrace{b_2(x_i)}_{} - \overbrace{\big(y_i - b_1(x_i)\big)}^{\text{For each object} \to \text{new label}}\Big)^2 \to \min_{b_2(x)}$$

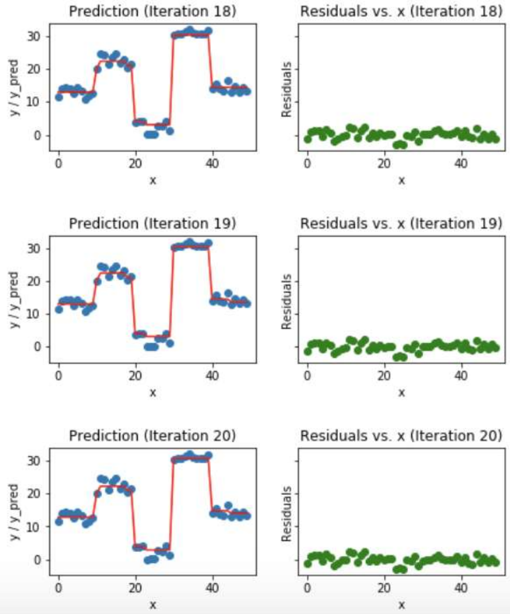
Who is b?

we don't care

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\Big(b_3(x_i) - \big(y_i - b_1(x_i) - b_2(x_i)\big)\Big)^2 \to \min_{b_3(x)}$$

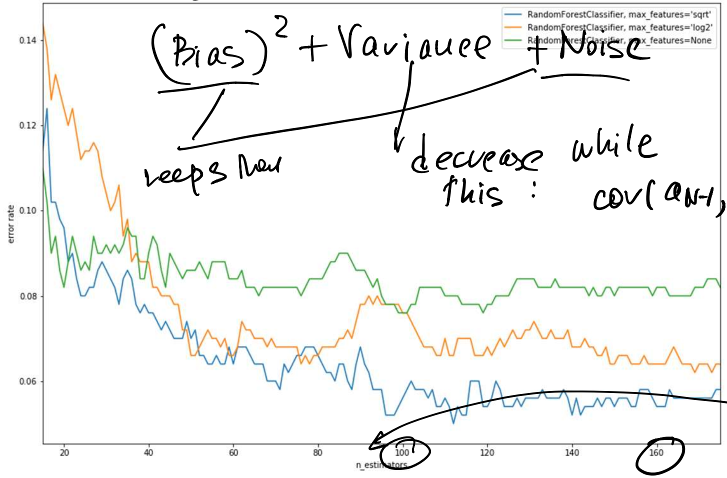

Scatter plot of x vs. y
y
x
0
10
20
30
40
50
0
5
10
15
20
25
30


Prediction (Iteration 1)
Residuals vs. x (Iteration 1)
Prediction (Iteration 2)
Residuals vs. x (Iteration 2)
Prediction (Iteration 3)
Residuals vs. x (Iteration 3)
y / y_pred
Residuals
x
value
keep
use no errors
Boosting
. low variance
. big bias
depth=1


Prediction (Iteration 18)
Residuals vs. x (Iteration 18)
Prediction (Iteration 19)
Residuals vs. x (Iteration 19)
Prediction (Iteration 20)
Residuals vs. x (Iteration 20)
y / y_pred
Residuals
x

Random Forest

RF $\Rightarrow$ Bagging

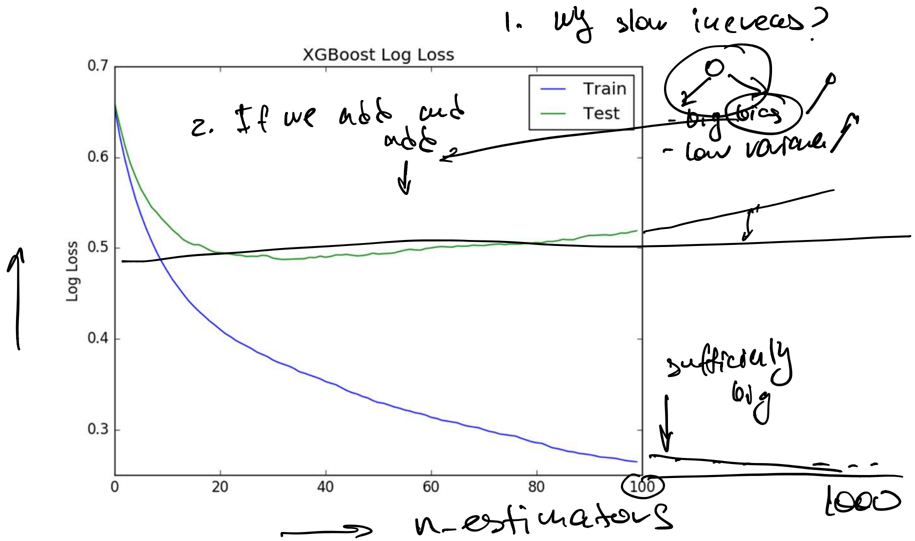


$(\text{Bias})^2 + \text{Variance} + \text{Noise}$

keeps then

decrease this : while $\text{cov}(a_{N-1}, b_N) \approx 0.$

161


1. why slow increases?
XGBoost Log Loss
2. If we add and add
- big bias
- low variance
Train
Test
Log Loss
0.7
0.6
0.5
0.4
0.3
0
20
40
60
80
100
sufficiently big
1000
n_estimators

Outlines

1) define new labels

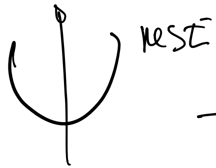
2) train b_n

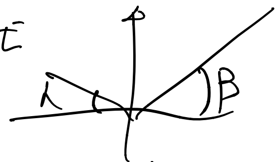
- 1) MSE loss ✓
- 2) Boosting can overfit.

(CV for hyperpar., ... ,)

$\Rightarrow$ MSE ✓

$\Rightarrow$? $\rightarrow$ classification ?





$\alpha < \beta$

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

- Learn N model:

$$\frac{1}{\ell} \sum_{i=1}^{\ell} L\Big(y_i, \underbrace{a_{N-1}}_{\text{fixed}}(x_i) + \underbrace{b_N}_{\text{optimize}}(x_i)\Big) \to \min_{b_N(x)}$$

Don't know
how to solve

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(y_i, a_{N-1}(x_i) + b_N(x_i)\big) \to \min_{b_N(x)}$$

MSE ✓

- for other?

Know this

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(y_i - a_{N-1}(x_i), b_N(x_i)\big) \to \min_{b_N(x)}$$

general loss

(as an MSE)

new estimator

- Regression problem ✓
  $y \in \mathbb{R}$ $y_1, y_2 - y_3$
- $y \in \{-1, 1\}$

$y_1 - y_2$?

-

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(y_i - a_{N-1}(x_i), b_N(x_i)\big) \to \min_{b_N(x)}$$

- b_N $y_i - a_{N-1}(x_i)$

i we ok.

$y_i - a_{N-1}$ $\cancel{\cdot\, b_N} = 0$

$+1 \quad +1 \quad = 0$

$-1 \quad -1 \quad = 0$

$-1 \neq +1 \quad = \pm 2.$

"okay"

classification rule

$$\boxed{a_N(x) = \operatorname{sign} \sum_{n=1}^{N} b_n(x)}$$

loss

$$L(y, z) = \log(1 + \exp(-yz))$$

- Logistic Regression

$$\frac{1}{\ell} \cancel{\sum_{i=1}^{\ell}} \log\Big(1 + \exp\Big(-\big(\underbrace{y_i - a_{N-1}(x_i)}_{\text{residual}}\big)\big(\underbrace{b_N(x_i)}_{\text{new algorithm}}\big)\Big)\Big) \to \min_{b_N(x)}$$

+/-1   +/-1

$$a_N(x) = \operatorname{sign} \sum_{n=1}^{N} b_n(x)$$

$$L(y, z) = \log(1 + \exp(-yz))$$

-

$$\frac{1}{\ell} \sum_{i=1}^{\ell} \log\left(1 + \exp\left(-\left(y_i - a_{N-1}(x_i)\right) b_N(x_i)\right)\right) \to \min_{b_N(x)}$$

$$a_N(x) = \operatorname{sign} \sum_{n=1}^{N} b_n(x)$$

$$L(y, z) = \log(1 + \exp(-yz))$$

- $$\frac{1}{\ell} \sum_{i=1}^{\ell} \log\left(1 + \exp\left(-\left(y_i - a_{N-1}(x_i)\right) b_N(x_i)\right)\right) \to \min_{b_N(x)}$$

- If $y_i = a_{N-1}(x_i)$, i $\Rightarrow$ we don't use this object for $b_N(x_i)$
- $\Rightarrow$ $b_N(x_i)$ could output anything on object i

Логистическая функция потерь

$$a_N(x) = \operatorname{sign} \sum_{n=1}^{N} b_n(x)$$

$$L(y, z) = \log(1 + \exp(-yz))$$

-

$$\frac{1}{\ell} \sum_{i=1}^{\ell} \log\Bigl(1 + \exp\Bigl(-\bigl(y_i - a_{N-1}(x_i)\bigr) b_N(x_i)\Bigr)\Bigr) \to \min_{b_N(x)}$$

- $y_i - a_{N-1}(x_i) = \pm 2$

Логистическая функция потерь

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\log\left(1+\exp\left(-\frac{y_i - a_{N-1}(x_i)}{2}b_N(x_i)\right)\right)\to\min_{b_N(x)}$$

- $y_i = a_{N-1}(x_i)$,
- $y_i \neq a_{N-1}(x_i)$,

Логистическая функция потерь

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\log\left(1+\exp\left(-\frac{y_i - a_{N-1}(x_i)}{2}b_N(x_i)\right)\right) \to \min_{b_N(x)}$$

- $y_i = +1, \sum_{n=1}^{N-1} b_n(x_i) = -0.5 \to \qquad b_N(x_i) > 0.5$
- $y_i = +1, \sum_{n=1}^{N-1} b_n(x_i) = -100 \to \qquad b_N(x_i) > 100$

Логистическая функция потерь

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\log\left(1+\exp\left(-\frac{y_i-a_{N-1}(x_i)}{2}b_N(x_i)\right)\right)\to\min_{b_N(x)}$$

- $y_i=+1, \sum_{n=1}^{N-1} b_n(x_i)=-0.5 \to$ надо $b_N(x_i)>0.5$
- $y_i=+1, \sum_{n=1}^{N-1} b_n(x_i)=-100 \to$ надо $b_N(x_i)>100$
-
-

$b_N(x)$

$x \cdot y \Rightarrow 0$ if $x=0$

$\log x + \log y \approx$ mimic MSE loss

(MSLE)

- Mean Squared Logarithmic Error (среднеквадратичная логарифмическая ошибка)

$$\underbrace{L(y, z) = (\log(z+1) - \log(y+1))^2}$$

MSLE

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

$$L(y, z) = (\log(z + 1) - \log(y + 1))^2$$

- $$\frac{1}{\ell} \sum_{i=1}^{\ell} (\log(b_N(x_i) + 1) - \log(y_i - a_{N-1}(x_i) + 1))^2 \to \min_{b_N(x)}$$

MSLE

$$a_N(x) = \sum_{n=1}^{N} b_n(x)$$

$$L(y, z) = (\log(z+1) - \log(y+1))^2$$

- Может, просто обучаться на остатки, как в MSE?

$$\frac{1}{\ell}\sum_{i=1}^{\ell}(\log(b_N(x_i) + 1) - \log(y_i - a_{N-1}(x_i) + 1))^2 \to \min_{b_N(x)}$$

-

MSLE

$$\frac{1}{\ell}\sum_{i=1}^{\ell}(\log(b_N(x_i)+1)-\log(y_i-a_{N-1}(x_i)+1))^2 \to \min_{b_N(x)}$$

$\boldsymbol{y_i}$	$\boldsymbol{a_{N-1}(x_i)}$	$\boldsymbol{b_N(x_i)}$		
1000	100	2	0.09	13.7
2	0	2	1.2	1.2

- N :

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(y_i, a_{N-1}(x_i) + b_N(x_i)\big) \to \min_{b_N(x)}$$

- N :

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(y_i, a_{N-1}(x_i) + b_N(x_i)\big) \to \min_{b_N(x)}$$

- $a_{N-1}(x_i)$,

- N

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(y_i, a_{N-1}(x_i) + b_N(x_i)\big) \to \min_{b_N(x)}$$

- $a_{N-1}(x_i)$,

-

GD

$\min_{x} f(x)$

Boosting

$\sqrt{MSE}$

$(y-\hat{y})^2$

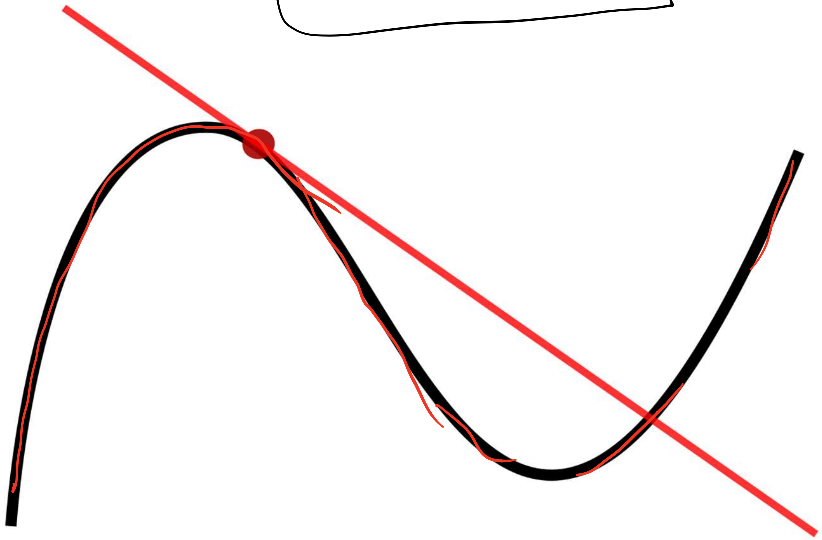
if we stay at x_0

$$f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{\alpha}{2} \|x - x_0\|_2^2$$

$\min_{x}$

$\Rightarrow$

$$x = x_0 - \frac{1}{\alpha} \nabla f(x_0)$$



current algorithm $\approx$ current point

- Learn Nth model

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L(y_i, a_{N-1}(x_i) + b_N(x_i)) \to \min_{b_N(x)}$$

$$\approx L(y_i, a_{N-1}) + \frac{\partial}{\partial z} L(y_i, a_{N-1})(\ldots)$$

$$\frac{1}{2}\, \| \quad \|$$

-

residual

$$s_i^{(N)} = -\frac{\partial}{\partial z} L(y_i, z)\Big|_{z=a_{N-1}(x_i)}$$

we don't speak about parametrization of b_N.

- N :

$$\frac{1}{\ell}\sum_{i=1}^{\ell} L\big(y_i, a_{N-1}(x_i) + b_N(x_i)\big) \to \min_{b_N(x)}$$

-

dataset

ensemble answers ≠ parameters

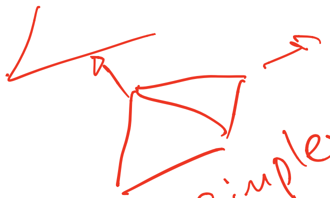
$$s_i^{(N)} = -\underbrace{\frac{\partial}{\partial z}}_{\text{Derivative}} L(y_i, z)\Big|_{z=a_{N-1}(x_i)}$$

current ensemble output

Задача обучения базовой модели

$$s_i^{(N)} = -\frac{\partial}{\partial z} L(y_i, z) \Big|_{z=a_{N-1}(x_i)}$$

$s_i^N > 0$, $s_i^N < 0$, $\cong 0$. $\Rightarrow$ direction of changing.



not change when no mistake

Градиентный бустинг

- N th :

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left(b_N(x_i) - s_i^{(N)}\right)^2 \to \min_{b_N(x)}$$

$$s_i^{(N)} = \underbrace{-\frac{\partial}{\partial z}L(y_i, z)}\Big|_{z=a_{N-1}(x_i)} -$$ residuals

Градиентный бустинг

- N : Tree, NN ∇_θ MSE

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left(b_N(x_i) - s_i^{(N)}\right)^2 \to \min_{b_N(x)}$$

$\in \mathbb{R}$

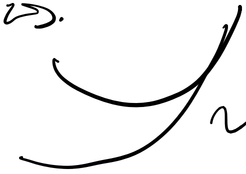
$$s_i^{(N)} = -\frac{\partial}{\partial z} L(y_i, z)\Big|_{z=a_{N-1}(x_i)} -$$

- $\boxed{\text{FIT}}$ base model b_N on x_i

should output $s_i^{(N)}$

Градиентный бустинг

GLM models
losses → cla, reg, poiss. convex

 ~ quad.

- fit Nth :

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left(b_N(x_i) - s_i^{(N)}\right)^2 \to \min_{b_N(x)}$$

$$s_i^{(N)} = -\frac{\partial}{\partial z} L(y_i, z)\Big|_{z=a_{N-1}(x_i)}$$ residual

- 1) GD in answers space
- 2) $\frac{\partial}{\partial z} L$ we use local information about loss function
- convex loss is locally quadratic MSE.

MSE

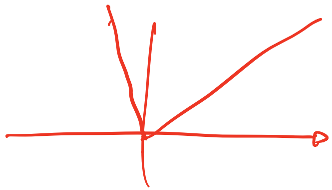
$$s_i^{(N)} = -\frac{\partial}{\partial z} L(y_i, z)\Big|_{z=a_{N-1}(x_i)} = -\frac{\partial}{\partial z}\frac{1}{2}(z - y_i)^2\Big|_{z=a_{N-1}(x_i)} =$$

$$= -(a_{N-1}(x_i) - y_i) = y_i - a_{N-1}(x_i)$$

$$s_i^{(N)} = -\frac{\partial}{\partial z} L(y_i, z)\Big|_{z=a_{N-1}(x_i)} = -\frac{\partial}{\partial z}\frac{1}{2}(z - y_i)^2\Big|_{z=a_{N-1}(x_i)} =$$

$$= -(a_{N-1}(x_i) - y_i) = y_i - a_{N-1}(x_i)$$

$$\underbrace{\frac{1}{\ell}\sum_{i=1}^{\ell}\Big(b_N(x_i) - \big(y_i - a_{N-1}(x_i)\big)\Big)^2 \to \min_{b_N(x)}}_{\text{(same update)}}$$



$$L(y,z)=\frac{1}{2}\left([z<y](z-y)^2+5[z\ge y](z-y)^2\right)$$

$$s_i^{(N)} = -\frac{\partial}{\partial z}L(y_i,z)\Big|_{z=a_{N-1}(x_i)} =$$

count on diff in loss.

$$= [z<y](y-z)+5[z\ge y](y-z)$$

y(c)   locally it is MSE

$$s_i^{(N)} = [z < y](y - z) + 5[z \geq y](y - z)$$

- $y_i = 10, a_{N-1}(x_i) = 5$: $s_i = 5$
- $y_i = 10, a_{N-1}(x_i) = 15$: $s_i = -25$

Logistic

$$s_i^{(N)} = -\frac{\partial}{\partial z} L(y_i, z)\bigg|_{z=a_{N-1}(x_i)} =$$

$$= -\frac{\partial}{\partial z} \log(1 + \exp(-y_i z))\bigg|_{z=a_{N-1}(x_i)} =$$

$$= \frac{y_i}{1 + \exp\big(y_i a_{N-1}(x_i)\big)}$$

$$s_i^{(N)} = -\frac{\partial}{\partial z} L(y_i, z)\Big|_{z=a_{N-1}(x_i)} =$$

$$= -\frac{\partial}{\partial z} \log(1 + \exp(-y_i z))\Big|_{z=a_{N-1}(x_i)} =$$

$$= \frac{y_i}{1 + \exp\big(y_i a_{N-1}(x_i)\big)}$$

$$\frac{1}{\ell} \sum_{i=1}^{\ell} \left(b_N(x_i) - \frac{y_i}{1 + \exp\big(y_i a_{N-1}(x_i)\big)} \right)^2 \to \min_{b_N(x)}$$

$$\frac{\partial}{\partial z} L$$

fit

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\Big(\underbrace{b_N(x_i)}_{\text{new algorithm}} - \underbrace{\frac{y_i}{1+\exp\big(y_i a_{N-1}(x_i)\big)}}_{\text{residuals}}\Big)^2 \to \min_{b_N(x)}$$

- $\underbrace{y_i a_{N-1}(x_i)}_{?} \gg 0$ $\qquad \underbrace{\frac{y_i}{1+\exp\big(y_i a_{N-1}(x_i)\big)}}_{\gg 0} \approx 0$

- $y_i a_{N-1}(x_i) \ll 0$ $\qquad \frac{y_i}{1+\exp\big(y_i a_{N-1}(x_i)\big)} \approx \pm 1$

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left(b_N(x_i) - \frac{y_i}{1+\exp\big(y_i a_{N-1}(x_i)\big)}\right)^2 \to \min_{b_N(x)}$$

- $$\frac{y_i}{1+\underbrace{\exp\big(y_i a_{N-1}(x_i)\big)}_{\to\,\infty}} \approx 0$$

- $$\frac{y_i}{1+\underbrace{\text{e}\quad\big(y_i a_{N-1}(x_i)\big)}_{\to\,0}} \approx \pm 1$$

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left(b_N(x_i) - \frac{y_i}{1 + \exp\big(y_i a_{N-1}(x_i)\big)}\right)^2 \to \min_{b_N(x)}$$

- $y_i = +1, a_{N-1}(x_i) = -0.7 \colon s_i = 0.67$
- $y_i = +1, a_{N-1}(x_i) = 2 \colon s_i = 0.12$

$$a_N(x) = a_{N-1}(x_i) + b_N(x_i)$$

- N :

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\left(b_N(x_i) - s_i^{(N)}\right)^2 \to \min_{b_N(x)}$$

- $s_i^{(N)} = -\frac{\partial}{\partial z} L(y_i, z)\Big|_{z=a_{N-1}(x_i)}$ —

u · ou ?

Hyper parameters

GD. α step-size

b_N

not limited with 1.

$$a_N(x) = a_{N-1}(x_i) + \eta b_N(x_i)$$

(library for boosting)

- $\eta \in (0, 1]$ — "learning rate" xgboost
-
-

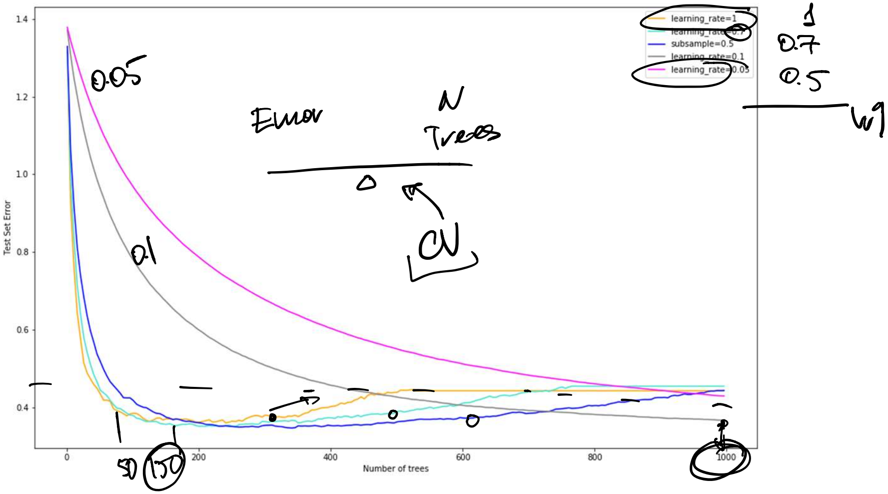
CV.

0 1

$\eta,>>$

1) less η more ensemble's members

b_N ← 2) big bias, low variance

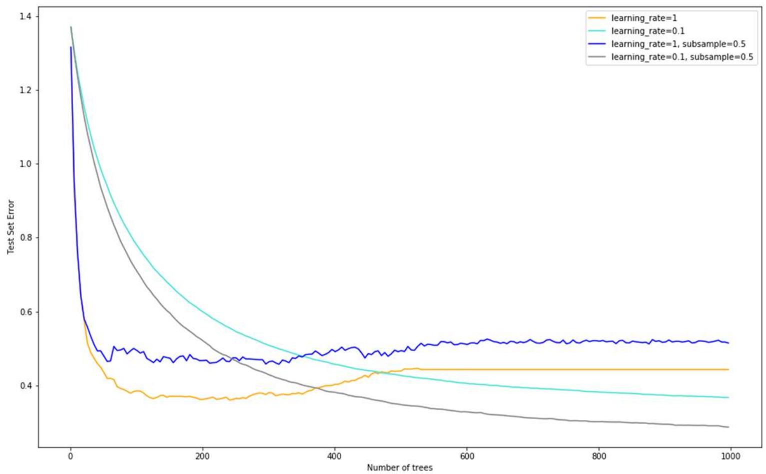

learning_rate=1
learning_rate=0.7
subsample=0.5
learning_rate=0.1
learning_rate=0.05
Test Set Error
Number of trees
1.4
1.2
1.0
0.8
0.6
0.4
0
200
400
600
800
1000
0.05
0.1
Error
N Trees
CV
50
150
1
0.7
0.5

Regularization

⇏ Randomization

↓ (Tree)

- feature space → restrict for $10 \ll 1000$ $\sqrt{K}$
- sample space


learning_rate=1
learning_rate=0.1
learning_rate=1, subsample=0.5
learning_rate=0.1, subsample=0.5
Test Set Error
Number of trees
1.4
1.2
1.0
0.8
0.6
0.4
0
200
400
600
800
1000

- https://www.gormanalysis.com/blog/gradient-boosting-explained/
- https://youtu.be/3CC4N4z3GJc
- https://en.wikipedia.org/wiki/Gradient_boosting
- https://dyakonov.org/2017/06/09/градиентный-бустинг/