

Методы машинного обучения в задаче предсказания погоды предварительные результаты

Лена Волжина

руководитель Е.Г. Михайлова

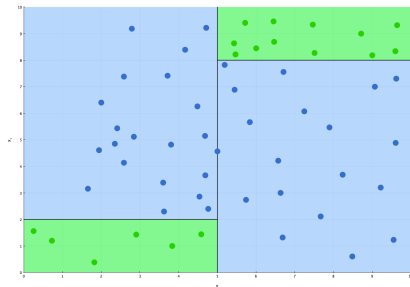
СПбГУ, мат-мех, кафедра ИАС

27 декабря 2017г

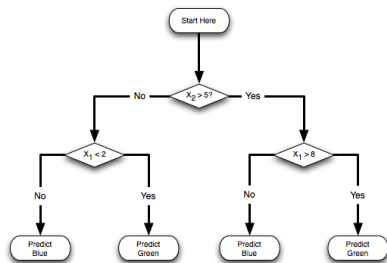
Введение

- ▶ Погода – состояние нижнего слоя атмосферы
- ▶ Температура, погодное явление, влажность, давление, ветер, ...
- ▶ Метеостанции, радиозонды, спутники
- ▶ Численный прогноз погоды
- ▶ Глобальные модели: GFS, JMA, ECMWF, CMC; региональная: WRF

Решающие деревья



(a) классификация наблюдений

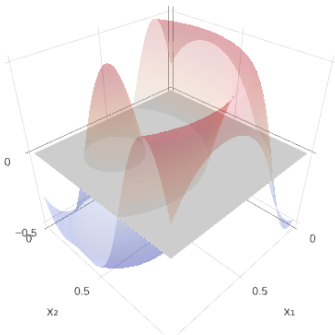


(b) решающее дерево

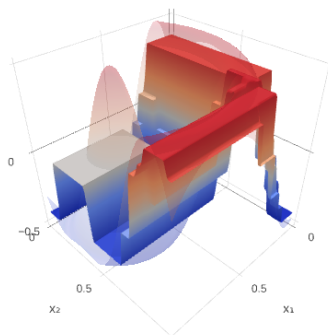
<https://alliance.seas.upenn.edu/~cis520/wiki/index.php?n=Lectures.DecisionTrees>

Градиентный бустинг

target function $f(\mathbf{x})$ and prediction of previous trees $D(\mathbf{x})$

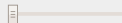
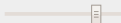


residual $R(\mathbf{x})$ and prediction of next tree $d_n(\mathbf{x})$



Tree depth: 5

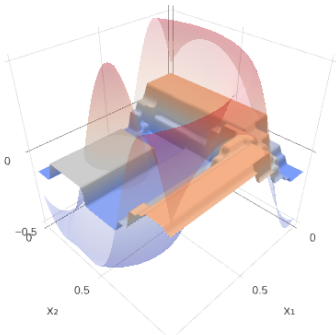
Number of built trees: 0



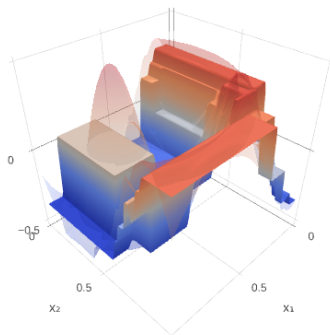
http://arogozhnikov.github.io/2016/06/24/gradient_boosting_explained.html

Градиентный бустинг

target function $f(x)$ and prediction of previous trees $D(x)$

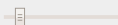
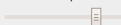


residual $R(x)$ and prediction of next tree $d_n(x)$



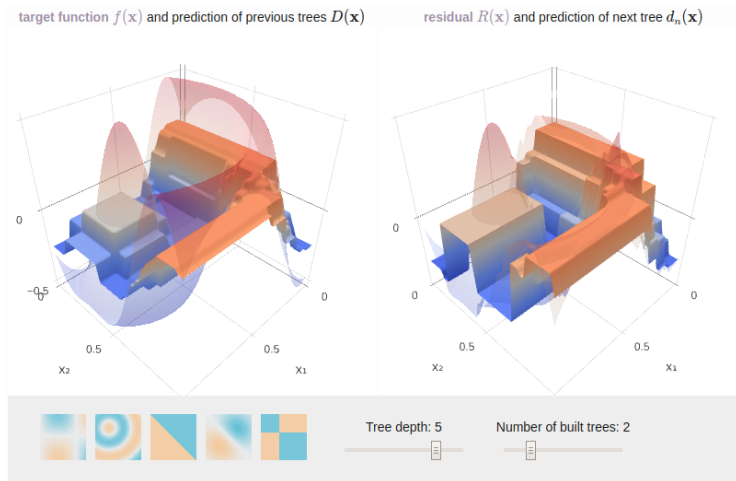
Tree depth: 5

Number of built trees: 1



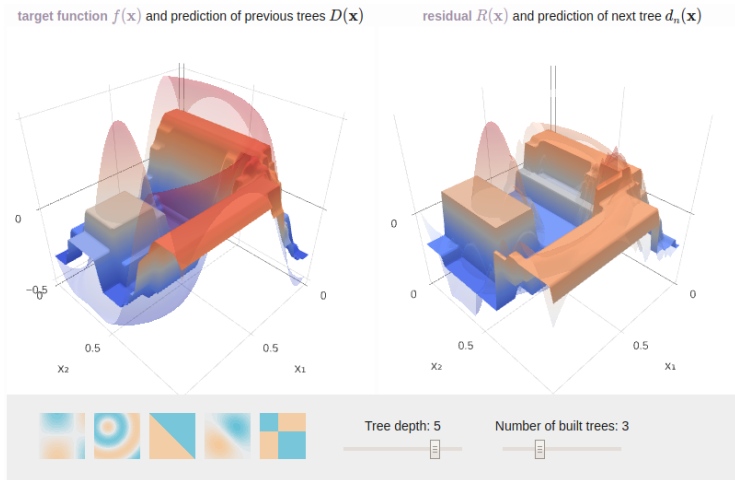
http://arogozhnikov.github.io/2016/06/24/gradient_boosting_explained.html

Градиентный бустинг



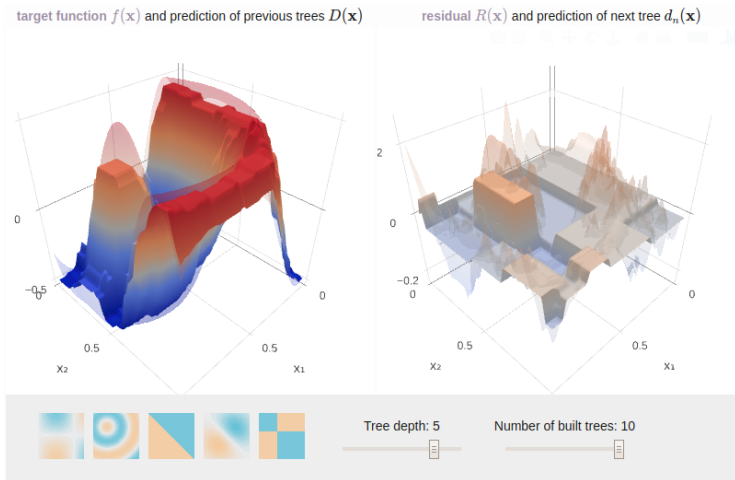
http://arogozhnikov.github.io/2016/06/24/gradient_boosting_explained.html

Градиентный бустинг



http://arogozhnikov.github.io/2016/06/24/gradient_boosting_explained.html

Градиентный бустинг



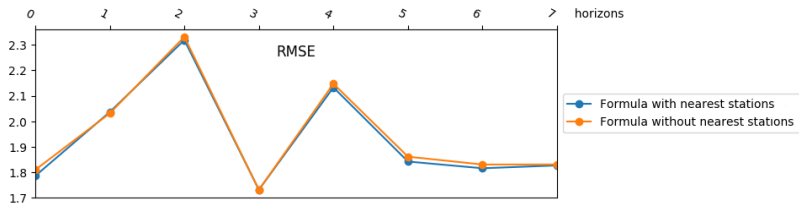
http://arogozhnikov.github.io/2016/06/24/gradient_boosting_explained.html

Эксперименты

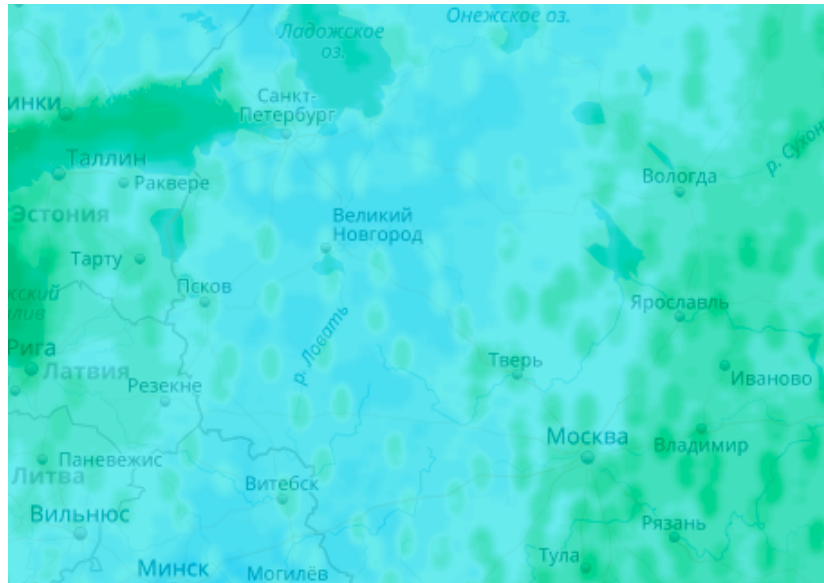
- ▶ $temperature_delta = temperature - climate_temperature$
- ▶ В момент $gentime$ прогнозируем погоду в момент $time$
- ▶ $X = \{x_i = (f_i^1, \dots, f_i^M), i \in \overline{1..N}\}, y = \{y_i, i \in \overline{1..N}\}$
- ▶ f_i : прогнозы поставщиков, климатические данные, вспомогательные переменные
- ▶ Задача: добавить показания ближайших станций в момент $gentime$ (на 7 часов вперёд)
- ▶ Метрика: $RMSE(y, \hat{y}) = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}$

Эксперименты: наивное решение

Модели, обученные на всех краткосрочных горизонтах



Эксперименты: только первые N часов

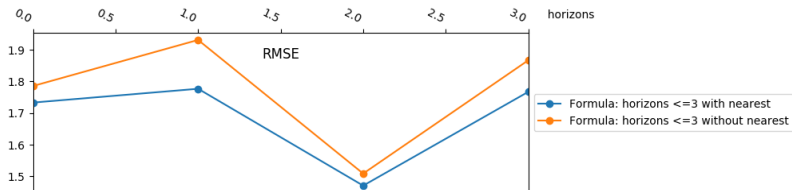


Эксперименты: сглаживание

$smoothed_fact_temperature_delta :=$

$$\begin{cases} fact \cdot \left(1 - \frac{dist}{max_distance}\right) + forecast \cdot \frac{dist}{max_distance}, & \text{if } dist \leq max_distance \\ forecast, & \text{if } dist > max_distance \end{cases},$$

где $fact$ – исходное значение с одной из ближайших станций,
 $dist$ – расстояние до неё, $forecast$ – прогноз поставщика



Дальнейшее развитие

- ▶ не только *temperature_delta*
- ▶ кросс-валидация
- ▶ параметры градиентного бустинга
- ▶ другие способы сглаживания

