

- линейные методы
- композиция деревьев — случайный лес
- композиция деревьев — градиентный бустинг

- регрессия, $Y = \mathbb{R} \Rightarrow$ линейная регрессия:

$$a(x) = \langle w, x \rangle$$

$$Q(w) = \sum_{i=1}^{\ell} (\langle w, x_i \rangle - y_i)^2 \rightarrow \min_w$$

- классификация, $Y = \{-1, +1\} \Rightarrow$ логистическая регрессия

$$a(x) = \text{sign}\langle w, x \rangle$$

$$Q(w) = \sum_{i=1}^{\ell} \log(\exp(-y_i(\langle w, x_i \rangle))) \rightarrow \min_w$$

- регуляризация:

$$Q(w) + \sum_{j=1}^d w_j^2 \rightarrow \min_w$$

$$Q(w) + \sum_{j=1}^d |w_j| \rightarrow \min_w$$

Преимущества:

- мало параметров, быстро обучаются
- могут настраиваться стохастическим градиентным спуском, не требуют хранения выборки в памяти
- интерпретируемые
- хорошо подходят для задач с малым числом объектов и большим числом признаков
- хорошо подходят для разреженных данных (dummy-кодирование, мешок слов)

Недостатки:

- низкая выразительность
- требуют предобработки данных

Средний прогноз решающих деревьев:

$$a(x) = \frac{1}{N} \sum_{n=1}^N b_n(x).$$

Обучение:

- деревья строятся независимо
- каждое дерево обучается по подвыборке объектов
- лучший признак для разбиения в вершине выбирается из подмножества признаков
- дерево строится до тех пор, пока в каждом листе не окажется по одному объекту
- отдельные деревья крайне переобучены; усреднение прогнозов устраняет проблему

Преимущества:

- сильный алгоритм, способен восстанавливать сложные зависимости
- не требует масштабирования признаков
- устойчив к шумам в признаках
- легко распараллелить: каждое дерево обучается независимо
- очень мало гиперпараметров
- нет переобучения при увеличении числа деревьев

Недостатки:

- деревья могут долго обучаться из-за большой глубины
- большой размер деревьев, долго строятся предсказания
- трудно настраивать на сложные меры качества
- плохо подходит для разреженных данных

Сумма прогнозов решающих деревьев:

$$a(x) = \sum_{n=1}^N \alpha_n b_n(x).$$

Обучение:

- деревья строятся последовательно
- деревья ограничиваются по глубине (2-6 уровней)
- каждое следующее дерево приближает градиент ошибки построенной композиции

Преимущества:

- сильный алгоритм, способен восстанавливать сложные зависимости
- не требует масштабирования признаков
- устойчив к шумам в признаках
- может настраиваться на любую дифференцируемую меру качества
- настраивается быстрее случайного леса

Недостатки:

- может переобучаться
- необходимо подбирать число деревьев
- плохо подходит для разреженных данных

- Линейные модели быстро настраиваются и подходят для разреженных данных
- Случайный лес подходит для небольших плотных выборок
- Градиентный бустинг подходит для больших плотных выборок, может оптимизировать любой дифференцируемый функционал