

Обучающая выборка: $X^\ell = (x_i, y_i)_{i=1}^\ell$, $x_i \in X$, $y_i \in \{-1, +1\}$

Базовые классификаторы: $b_1(x), \dots, b_T(x)$, $b_t: X \rightarrow \{-1, +1\}$

Простое голосование базовых классификаторов:

$$a(x) = \text{sign} \sum_{t=1}^T b_t(x)$$

Композиция $a(x)$ может быть лучше базовых $b_1(x), \dots, b_T(x)$, если они лучше случайного гадания и достаточно различны.

Способы повышения различности базовых классификаторов:

- обучение по случайным подвыборкам,
- обучение по выборке со случайными весами объектов,
- обучение по случайным подмножествам признаков,
- использование различных моделей классификации,
- использование различных начальных приближений,
- использование рандомизации при обучении b_1, \dots, b_T .

Бэггинг (bagging, bootstrap aggregating) [Breiman, 1996]:
 $b_t(x)$ обучаются независимо по случайным подвыборкам
длины ℓ с повторениями (как в методе bootstrap),
доля объектов, попадающих в выборку: $(1 - \frac{1}{e}) \approx 0.632$

Метод случайных подпространств
(RSM, random subspace method) [Ho, 1998]:
 $b_t(x)$ обучаются по случайным подмножествам n' признаков.

Совместим обе идеи в одном алгоритме.

$\mathcal{F} = \{f_1, \dots, f_n\}$ — признаки,
 $\mu(\mathcal{G}, U)$ — метод обучения алгоритма по подвыборке $U \subseteq X^\ell$,
использующий только признаки из $\mathcal{G} \subseteq \mathcal{F}$.

Breiman L. Bagging predictors // Machine Learning, 1996.

Ho T. K. The random subspace method for constructing decision forests // IEEE Trans. on Pattern Analysis and Machine Intelligence, 1998.

Вход: обучающая выборка X^ℓ ; параметры: T

ℓ' — длина обучающих подвыборок;

n' — длина признакового подописания;

ε_1 — порог качества базовых алгоритмов на обучении;

ε_2 — порог качества базовых алгоритмов на контроле;

Выход: базовые алгоритмы b_t , $t = 1, \dots, T$;

1: **для всех** $t = 1, \dots, T$

2: $U_t :=$ случайное подмножество X^ℓ длины ℓ' ;

3: $\mathcal{G}_t :=$ случайное подмножество \mathcal{F} длины n' ;

4: $b_t := \mu(\mathcal{G}_t, U_t)$;

5: **если** $Q(b_t, U_t) > \varepsilon_1$ или $Q(b_t, X^\ell \setminus U_t) > \varepsilon_2$ **то**

6: не включать b_t в композицию;

Композиция — простое голосование: $a(x) = \text{sign} \sum_{t=1}^T b_t(x)$.

Обучение случайного леса:

- бэггинг над решающими деревьями
- усечение дерева (pruning) не производится
- признак в каждой вершине дерева выбирается из случайного подмножества k из n признаков
- для регрессии рекомендуется $k = \lfloor n/3 \rfloor$
- для классификации рекомендуется $k = \lfloor \sqrt{n} \rfloor$

Подбор числа деревьев T по критерию *out-of-bag*:

число ошибок на объектах x_i , если не учитывать голоса деревьев, для которых x_i был обучающим:

$$\text{out-of-bag}(a) = \sum_{i=1}^{\ell} \left[\text{sign} \left(\sum_{t=1}^T [x_i \notin U_t] b_t(x_i) \right) \neq y_i \right] \rightarrow \min$$

Это несмещённая оценка обобщающей способности.

Breiman L. Random Forests // Machine Learning. 2001.

- RF — один из самых сильных методов машинного обучения.
- RF обычно лишь немного уступает градиентному бустингу.
- Бэггинг позволяет вычислять оценки out-of-bag:
 - для оптимизации числа базовых алгоритмов T ,
 - для оценивания важности признаков,
 - для выбора параметра k в RF.
- RSM хорош, когда много неинформативных признаков.
- Бэггинг, RF, RSM эффективно распараллеливаются.