

XGBoost: Мощный фреймворк машинного обучения

XGBoost, или Extreme Gradient Boosting, является мощным алгоритмом машинного обучения, используемым для решения задач классификации и регрессии.

Соколов С.А., Корьев М.А. гр 5030102/10101



Описание алгоритма XGBoost

Градиентный бустинг

Последовательное обучение деревьев решений для прогнозирования.

Формирование ансамбля моделей для повышения точности.

Регуляризация

Предотвращение переобучения модели за счет штрафных функций.

Повышение обобщающей способности, улучшая прогнозирование на новых данных.



Принцип градиентного бустинга

1

1. Базовая модель

Обучение первого дерева
решений

2

2. Остаточная
ошибка

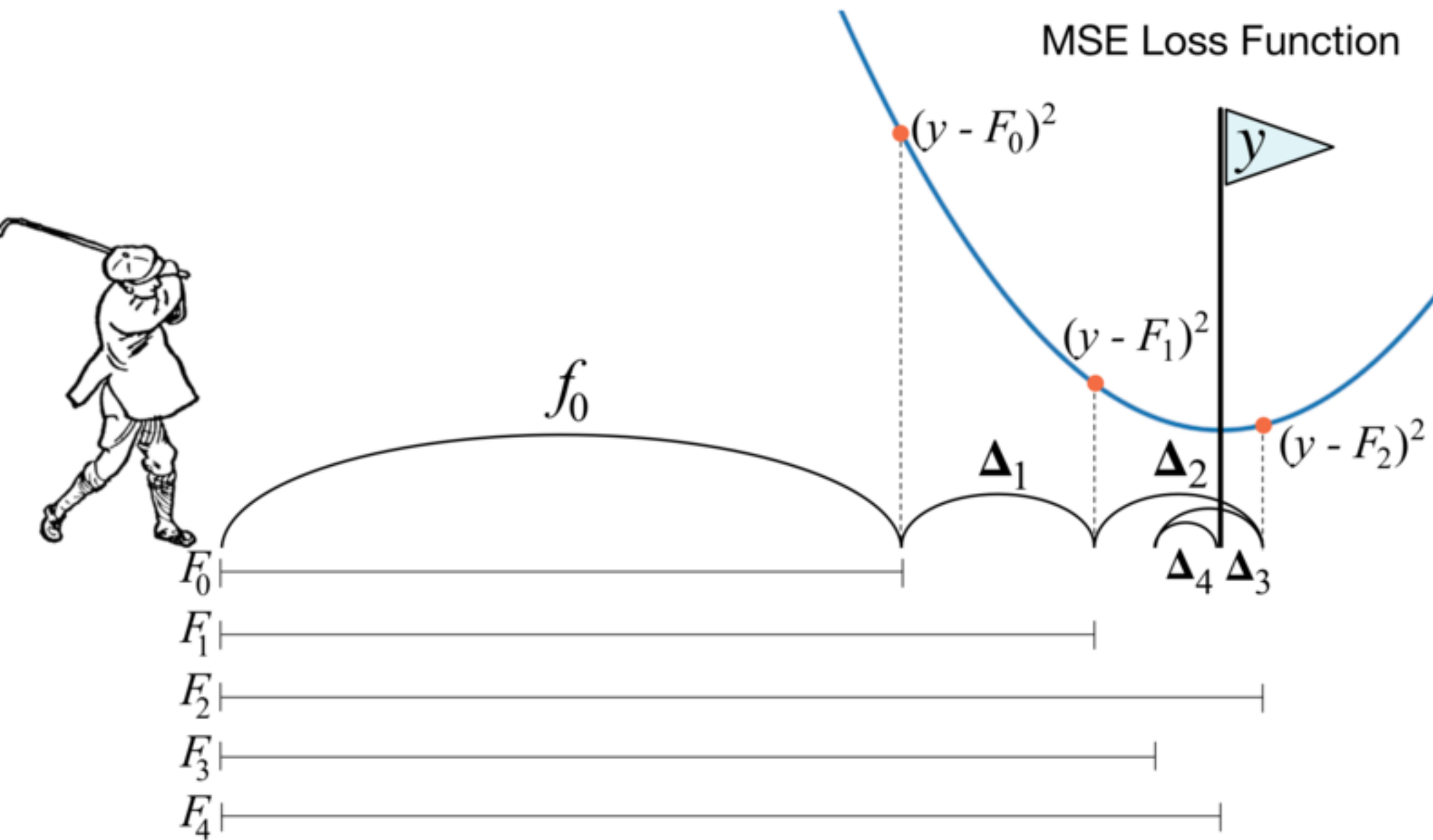
Анализ ошибок базовой
модели

3

3. Обучение следующего дерева

Обучение нового дерева, минимизирующего остаточную
ошибку





$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$ — функция для оптимизации градиентного бустинга

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + 0.5 h_i f_t^2(x_i)) + \Omega(f_t)$$



Обучение ансамбля моделей

Последовательное обучение деревьев

XGBoost строит ансамбль, последовательно добавляя новые деревья решений. Каждое последующее дерево учитывает ошибки предыдущих.

Отличие от бэггинга

В отличие от бэггинга, где деревья обучаются независимо, в XGBoost они обучаются последовательно, корректируя ошибки друг друга.

Пример использования XGBoost

```
import xgboost as xgb
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split

iris = load_iris()
X_train, X_test, y_train, y_test = train_test_split(
    iris.data, iris.target, test_size=0.2, random_state=42
)

model = xgb.XGBClassifier()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

# Оценка производительности модели
```





Оценка производительности XGBoost

Метрики оценки

Точность, полнота, F1-мера,
AUC

Перекрестная проверка

Разделение данных на
обучающие и тестовые
подмножества

Гиперпараметры

Настройка параметров модели для достижения оптимальных
результатов



Особенности XGBoost



Высокая скорость

Оптимизированный алгоритм



Высокая точность

Превосходная точность
прогнозирования



Гибкость

Поддерживает различные типы
данных и задач

Реальные примеры применения

1

Финансовое моделирование

2

Распознавание образов

3

Анализ текста



Выводы

XGBoost — мощный фреймворк, предоставляющий современный подход к решению задач машинного обучения.

I ew w il·ion



Источники

1. "XGBoost." *NEERC Wiki*, neerc.ifmo.ru/wiki/index.php?title=XGBoost. Дата обращения: 6 декабря 2024.
2. Python Package Index PYPI: xgboost. Дата обращения: 6 декабря 2024.
3. XGBoost - ML winning solutions (incomplete list). Дата обращения: 6 декабря 2024.
4. Tree Boosting With XGBoost - Why Does XGBoost Win "Every" Machine Learning Competition? Дата обращения: 6 декабря 2024.
5. *Sagi, Omer; Rokach, Lior (2021). "Approximating XGBoost with an interpretable decision tree". Information Sciences. 572 (2021): 522-542. doi:10.1016/j.ins.2021.05.055.*
6. *Gandhi, Rohith Gradient Boosting and XGBoost (англ.). Medium (24 мая 2019). Дата обращения: 6 декабря 2024.*