

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования

Национальный исследовательский ядерный университет «МИФИ»

Отчет о научно-исследовательской работе
на тему:

Применение машинного обучения для классификации
событий в эксперименте DEAP-3600

Выполнил:
Студент гр. М19-115
Ильясов А.И.
Научный руководитель:
к.ф.-м.н.
Гробов А.В.

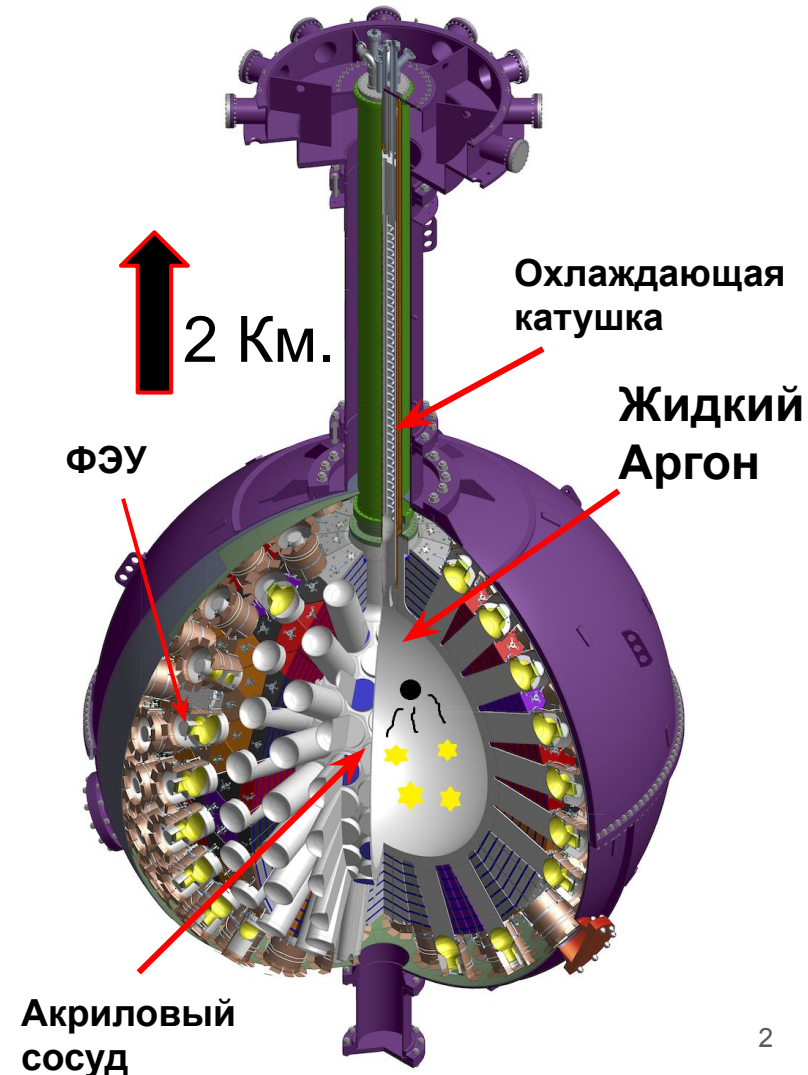
Москва, 2019



Детектор DEAP-3600

The DEAP Collaboration, Design and Construction of the DEAP-3600 Dark Matter Detector, Astropart. Phys. 108, 1 (2019).

- Расположен в 2 км. под землей;
- (3279 ± 96) кг. LAr (84 K) в UVA акриловом сосуде толщиной 5 см. и диаметром 1.7 м.;
- События регистрируются 255 ФЭУ;
- Шейный отдел детектора, который используется для поддержания нужной температуры детектора;
- Имеется нейтронная и черенковская мюонная защита;
- Детектор заключен в оболочку из нержавеющей стали и погружен в бассейн 300 т. ультрачистой воды



Фоновые события

Search for dark matter with a 231-day exposure of liquid argon using DEAP-3600 at SNOLAB (Arxiv: 1902.04048)

- Нейтроны
 - Космогенные
 - Радиогенные
- β и γ -излучение
 - События с электронной отдачей в рабочем объеме
 - Излучение Черенкова
- α -распады
 - В рабочем объеме
 - На внутренней поверхности акрилового сосуда
 - В шейном отделе детектора

$$N_{bkg}^{ROI} = N_{ER}^{ROI} + N_{Cher}^{ROI} + N_{n,rdg}^{ROI} + N_{n,csq}^{ROI} + N_{\alpha,AV}^{ROI} + N_{\alpha,neck}^{ROI}$$

Фоновые события

Search for dark matter with a 231-day exposure of liquid argon using DEAP-3600 at SNOLAB (Arxiv: 1902.04048)

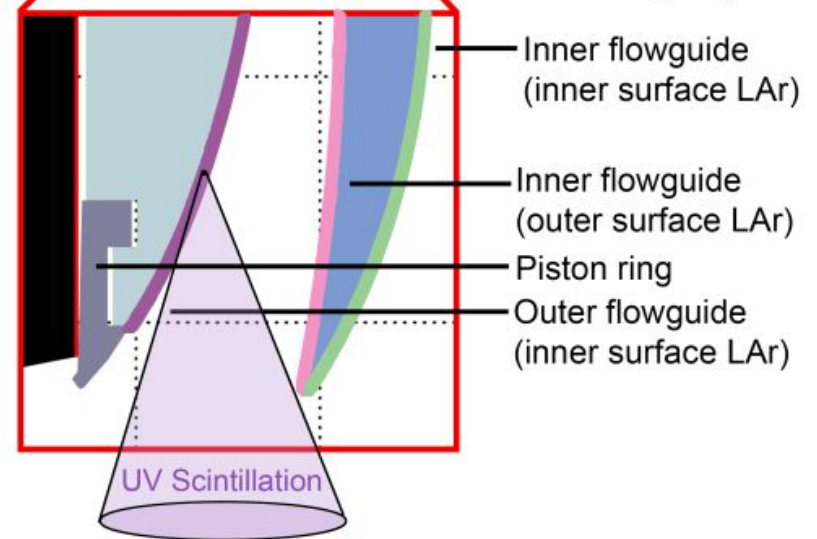
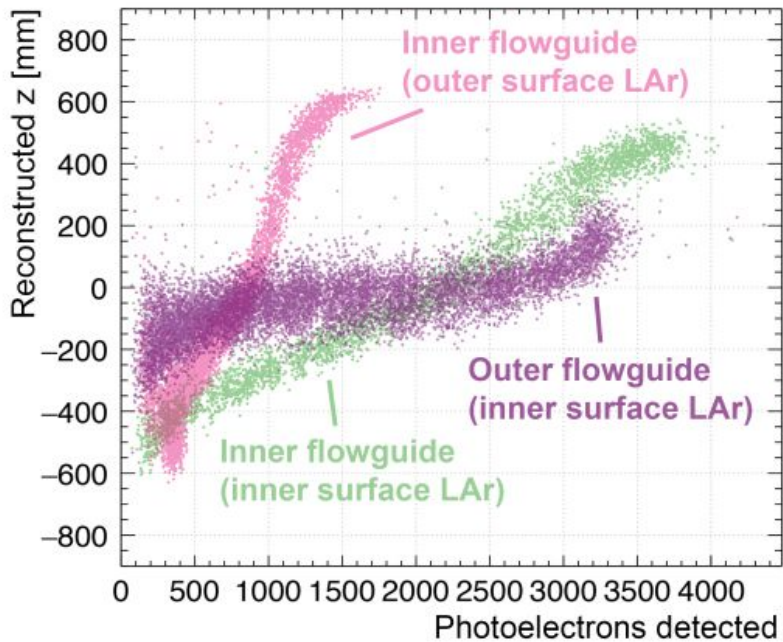
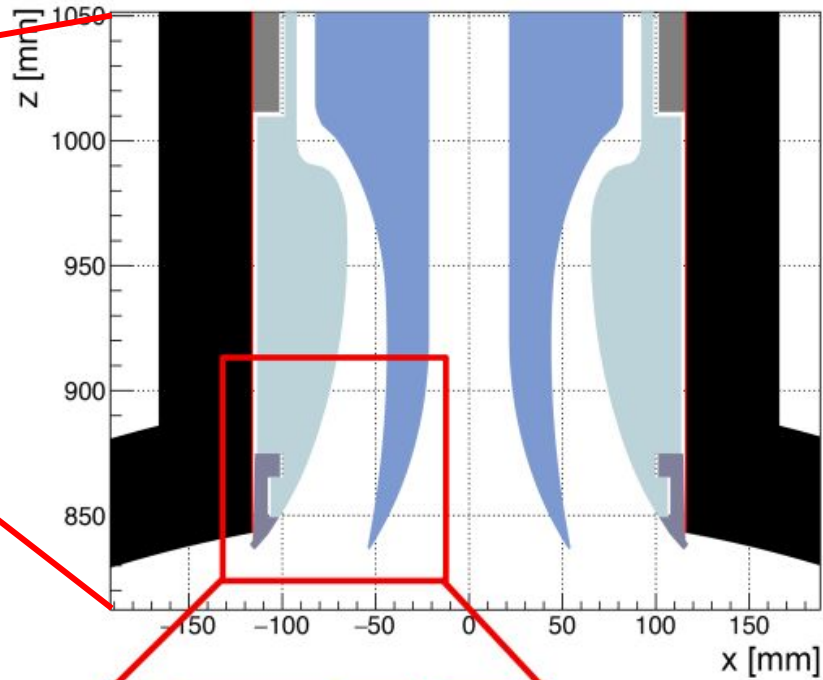
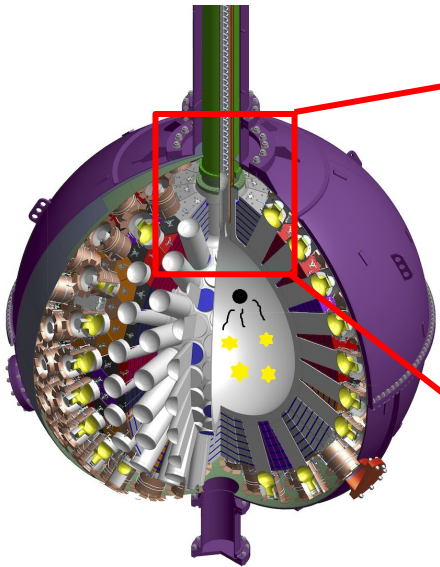
- Нейтроны
 - Космогенные
 - Радиогенные
 } Другие значения PE
 } Другая позиция

- β и γ-излучение
 - События с электронной отдачей в рабочем объеме
 - Излучение Черенкова
 } Материалы + PSD

- α-распады
 - В рабочем объеме
 - На внутренней поверхности акрилового сосуда
 - **В шейном отделе детектора**
 } PSD
 } другие значения PE
 } fiducial cut

$$N_{bkg}^{ROI} = N_{ER}^{ROI} + N_{Cher}^{ROI} + N_{n,rdg}^{ROI} + N_{n,csg}^{ROI} + N_{\alpha,AV}^{ROI} + N_{\alpha,neck}^{ROI}$$

α - распады



Критерии для шейных альфа распадов

- Предел на верхнее значение параметра PSD:

$$F_{prompt} \equiv \frac{\sum_{i|t_i \in (-28ns, 60ns)} Q_i}{\sum_{i|t_i \in (-28ns, 10\mu s)} Q_i}$$

- Ранние импульсы в ФЭУ газовой среды: Событие отбрасывается, если хотя бы 1 из первых **трех** фотоэлектронов регистрируется в ФЭУ газовой среды во время записи события.

Достигается уровень фоновых шейных альфа-распадов < 0.5 событий

НО

Уровень принятия сигнала опускается на 49%

Классификатор на основе деревьев решений

Структура:

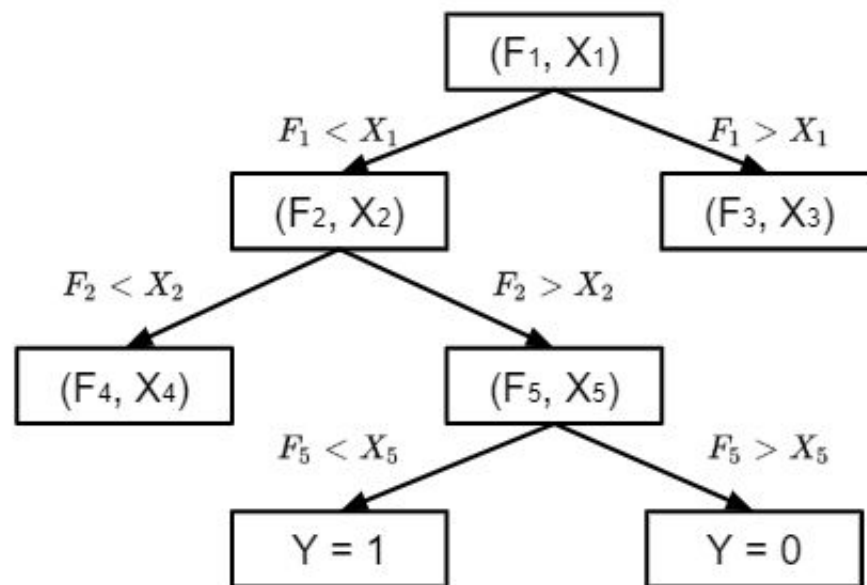
- Древоподобная структура;
- Каждый узел - "тест" на параметр;
- Каждая ветвь - результат теста.

Преимущества:

- Простота в понимании;
- Не требует большого количества данных.

Недостатки:

- Работает нестабильно с большим количеством данных.



Классификатор на основе деревьев решений

Структура:

- Древоподобная структура;
- Каждый узел - "тест" на параметр;
- Каждая ветвь - результат теста.

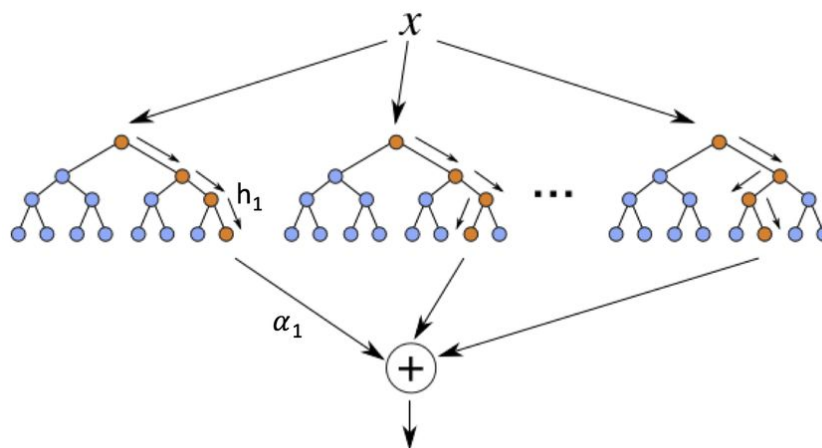
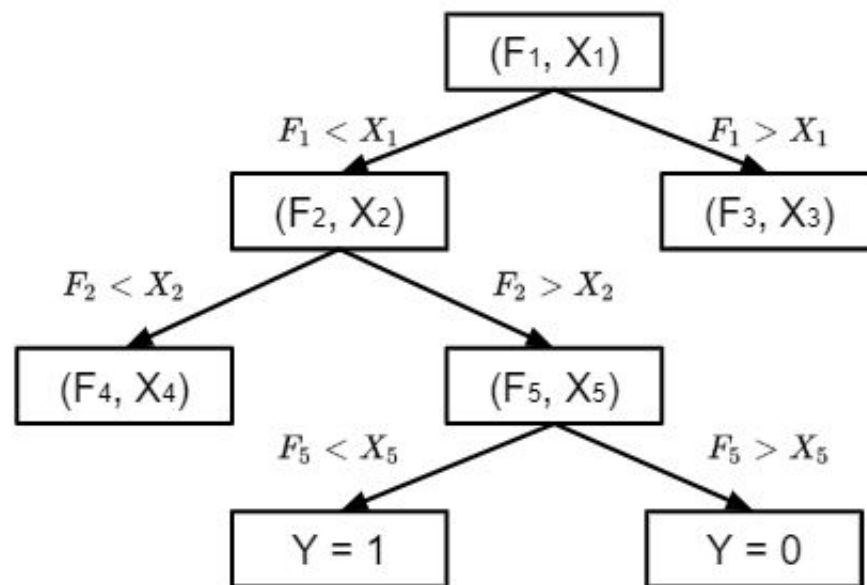
Преимущества:

- Простота в понимании;
- Не требует большого количества данных.

Недостатки:

- Работает нестабильно с большим количеством данных.

Решение



Метрики оценки качества классификатора

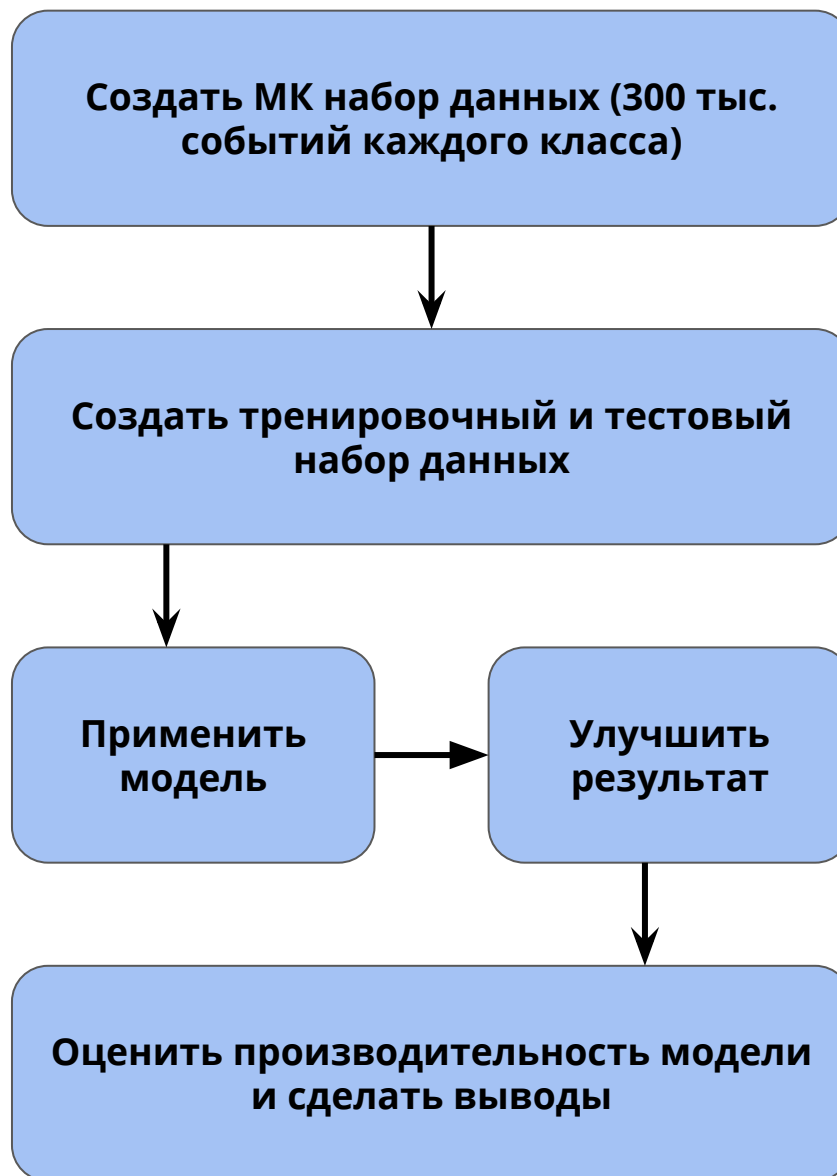
- Монте-Карло данные для ^{40}Ar и α -распадов
- Язык Python, пакет XGBoost
- Техника feature selection для выбора параметров
- 70% данных для тренировки и 30% для тестирования
- Использование поиска гиперпараметров по сетке
- Метрики качества и ROC-кривая - зависимость уровня отклонения фоновых событий от принятия сигнальных событий

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN}$$

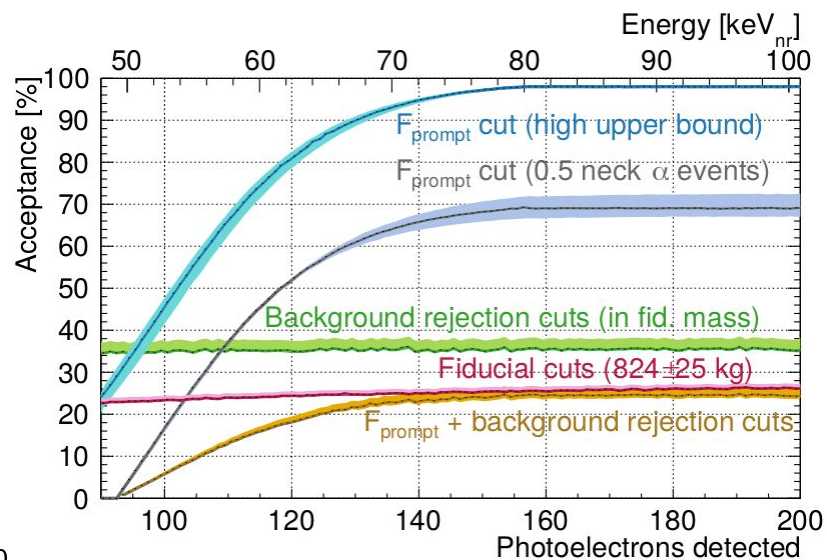
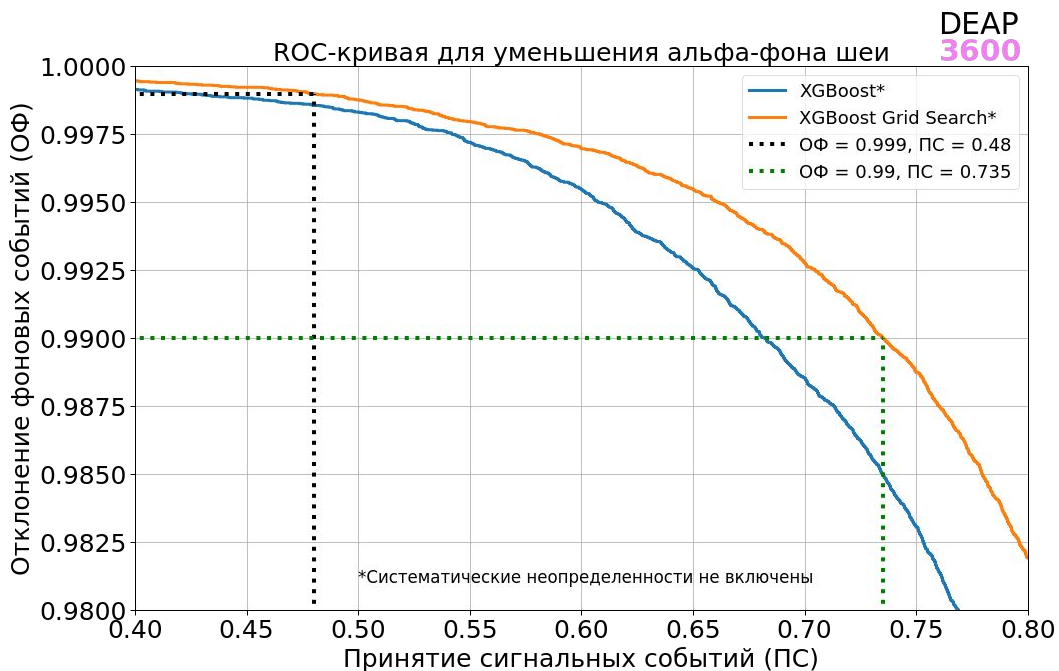
Параметры, необходимые для вычисления метрик		Истинная метка	
		1	0
Предсказанная метка	1	TP	FP
	0	FN	TN

Ход работы

План



Результат



XGBoost ROC-кривая - проверка работы классификатора без оптимизации

XGBoost Grid Search ROC-кривая - результат работы классификатора после подбора гиперпараметров с помощью поиска по сетке

	precision	recall	f1-score	support
0	0.92	0.91	0.92	62701
1	0.94	0.95	0.94	90623
micro avg	0.93	0.93	0.93	153324
macro avg	0.93	0.93	0.93	153324
weighted avg	0.93	0.93	0.93	153324

Заключение

- Использовали 300 тыс. событий Монте-Карло каждого класса;
- Имеем $\approx 93\%$ точности работы классификатора;
- На уровне подавления фона 99,9% мы имеем $\approx 48\%$ принятия сигнальных событий;
- Уже сейчас мы имеем результат, который показывает значительное улучшение разделения фоновых и сигнальных событий в шее детектора. Систематические неопределенности не были включены.

Планы

- Увеличение Монте-Карло статистики
- Применение дополнительных техник для увеличения принятия сигнальных событий до 90% на уровне отклонения фоновых 99.9%
- Объединение данного метода с другими (нейронная сеть, случайный лес и т.д.)

Градиентный бустинг над деревьями решений

$$y \approx \hat{f}(x),$$

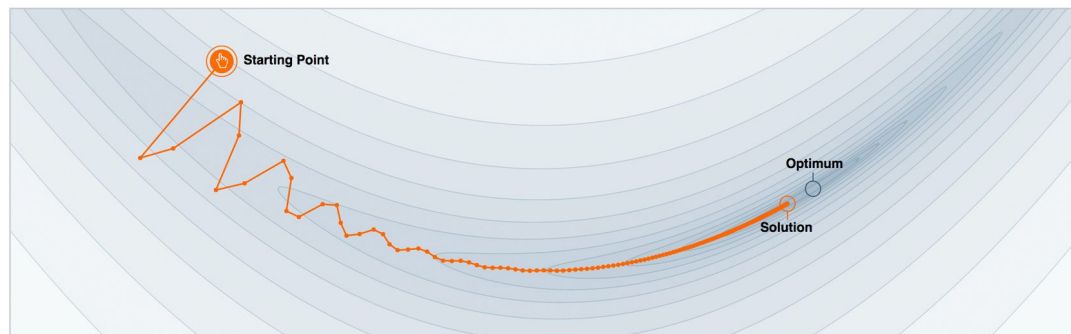
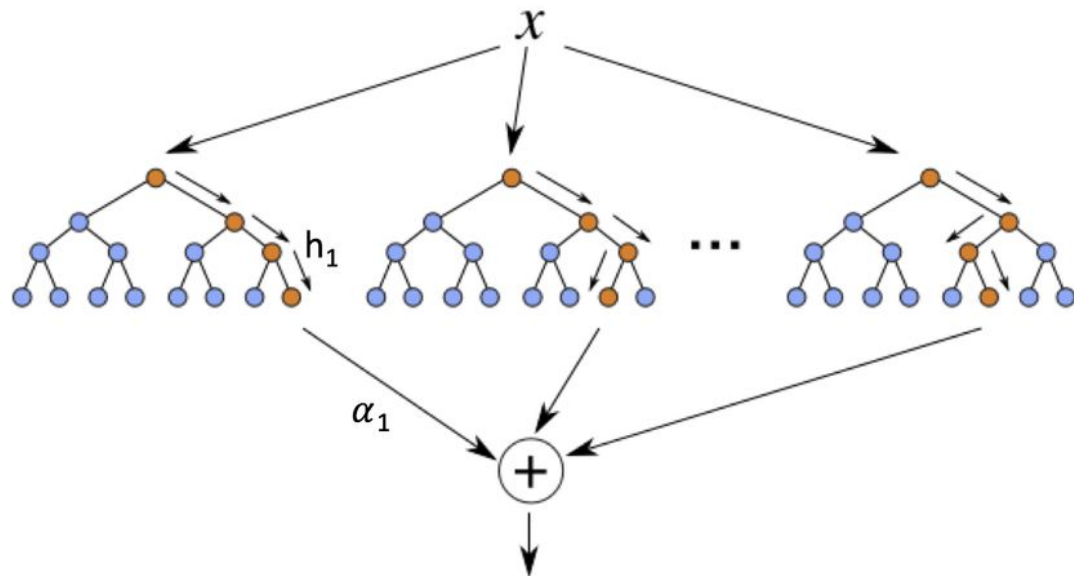
$$\hat{f}(x) = \arg \min_{f(x)} L(y, f(x))$$

Где:

x - входные переменные,

y - целевые переменные,

L - функция потерь



Randomized Search/GridSearch

Randomized Search:

- 1) Задается набор значений для каждого интересующего параметра
- 2) Задается количество проверок M
- 3) Задается метрика проверки
- 4) Проверяется M классификаторов с параметрами, случайно взятыми из набора и выдаются параметры классификатора с наилучшим результатом по оцениваемой метрике

Grid Search:

- 1) Задается пространство гиперпараметров
- 2) Задается метрика проверки
- 3) Проверяются любые возможные комбинации параметров и выдаются параметры классификатора с наилучшим результатом по оцениваемой метрике

