



# Применение методов машинного обучения для идентификации струй, образованных $W$ -бозоном

Студент: Ван Алина Маошэновна, М23-112

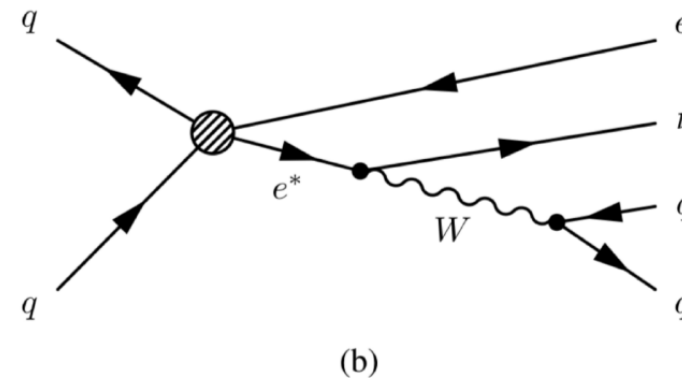
Научный руководитель: Мягков Алексей Григорьевич, к.ф.-м.н.



# Мотивация

## Проблемы Стандартной модели:

- Скрытая масса
- Проблема иерархии масс и структуры поколений
- Темная энергия и т.д.



Пример: Поиск возбужденного лептона с последующим распадом через калибровочный бозон

Идентификация толстых струй, образованных  $W$ -бозонами, распавшимися по адронной моде, является одним из главных составляющих этапов анализа данных с экспериментов по поиску новой физики.

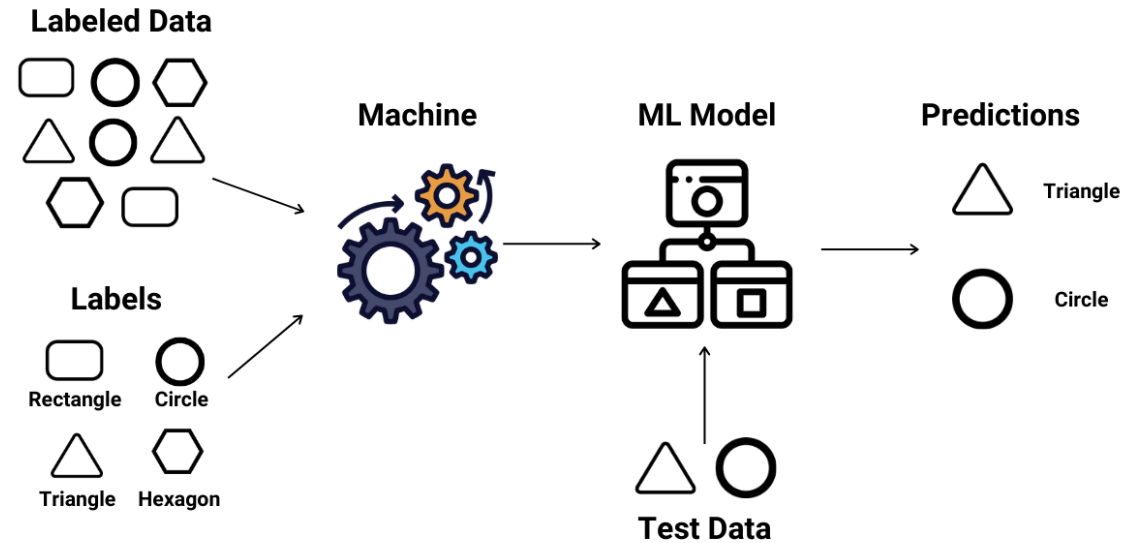
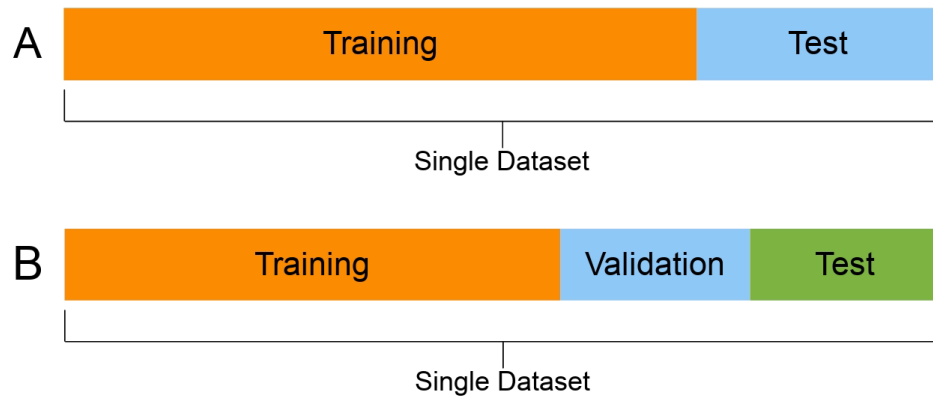
**Целью работы** является использование различных моделей глубокого обучения для решения задачи идентификации толстых струй, образованных  $W$ -бозоном.

В соответствии с поставленной целью задачами данной работы были:

- Ознакомление с различными моделями глубокого обучения для решения задачи бинарной классификации;
- Выбор дискриминирующих переменных (признаков) для объектов и формирование сигнального и фоновых деревьев для данной задачи
- Ознакомление с библиотекой глубокого обучения Keras;
- Обучение и тестирование моделей;
- Резюме результатов.

# Машинное обучение с учителем

**Машинное обучение** – это область прикладной математики, изучающая методы решения задач с использованием обучающих данных.



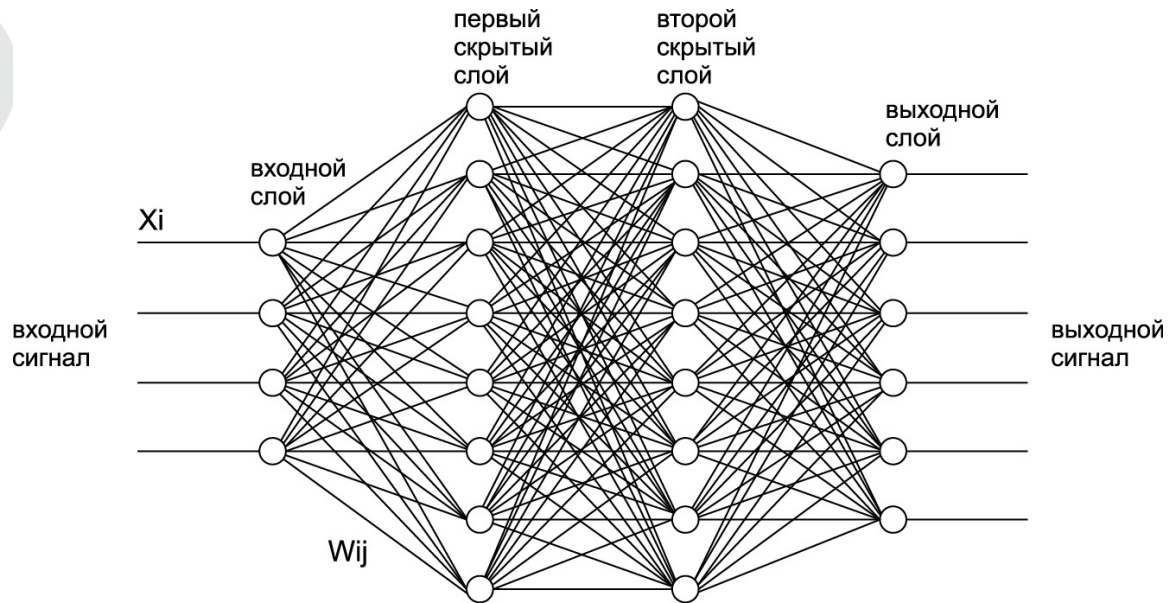
**Обучающая выборка** (на которой модель обучают)

**Валидационная выборка** (для оценки переобученности модели, для оценки ошибки прогнозирования при выборе модели, для настройки гиперпараметров и выбора лучшей модели), показывает, как может повести себя модель с новыми данными.

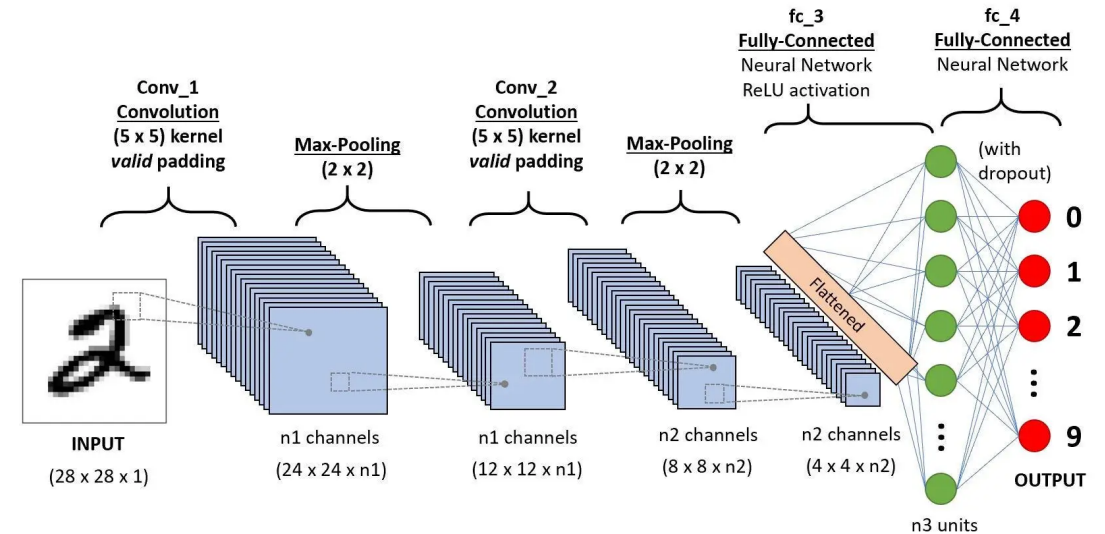
**Тестовая выборка** (для оценки работы готовой модели)

# Используемые архитектуры сетей

## Многослойный перцептрон (MLP)



## Сверточная нейронная сеть (CNN)



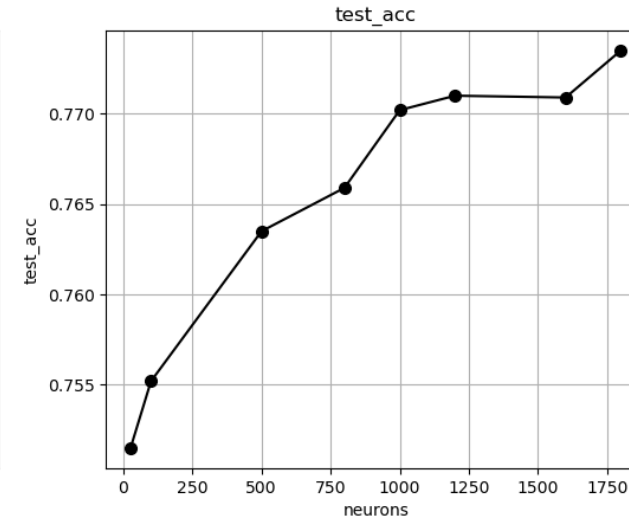
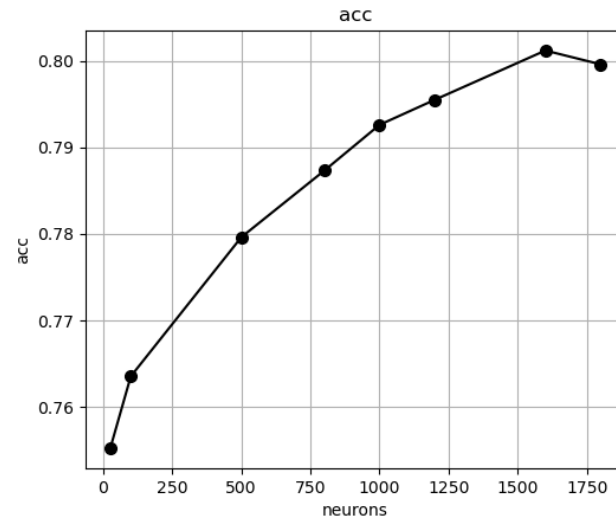
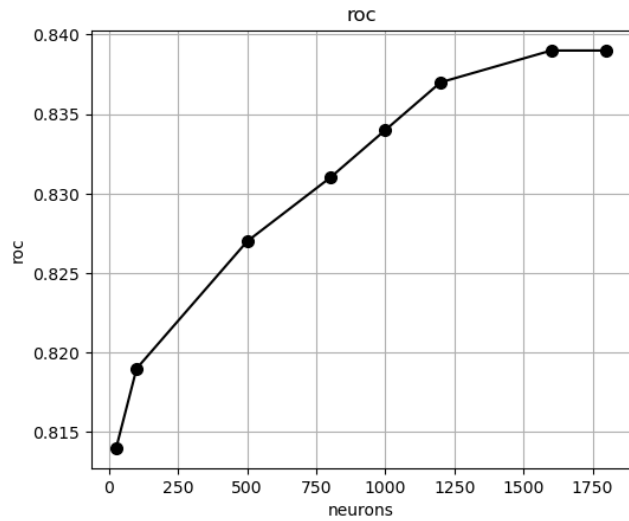
### Основные гиперпараметры MLP:

- размер батча
- скорость обучения
- функция активации
- количество нейронов
- количество слоев
- выбор алгоритма оптимизации

### Основные гиперпараметры CNN:

- те же, что и в MLP
- размер окна
- количество фильтров
- размерность фильтра

# Процесс работы. MLP(1)

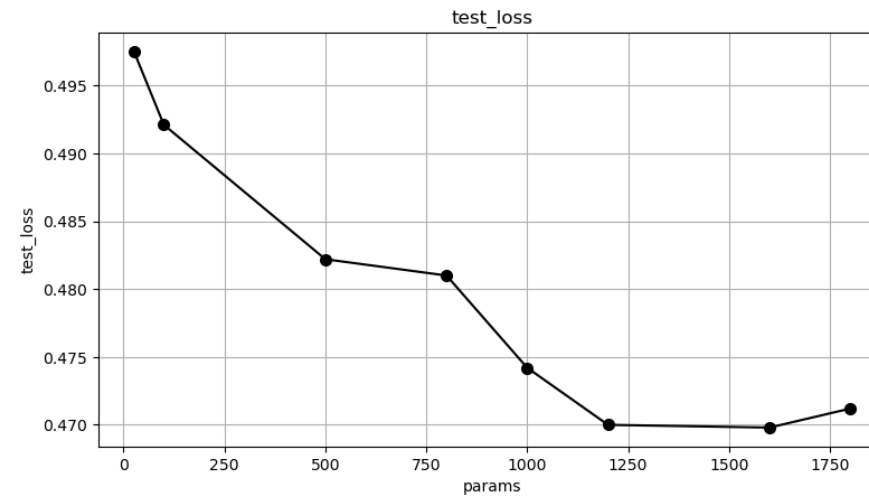
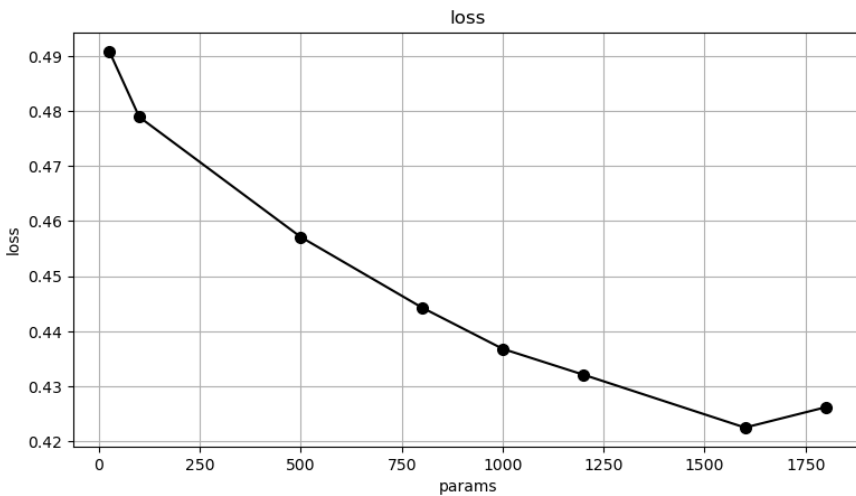


Однослойная  
сеть

Функция активации - ReLU

Batch = 512

LearningRate= 0.01

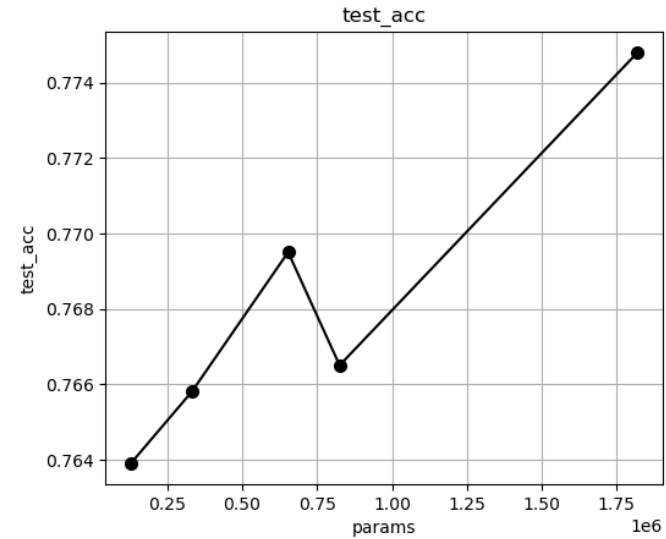
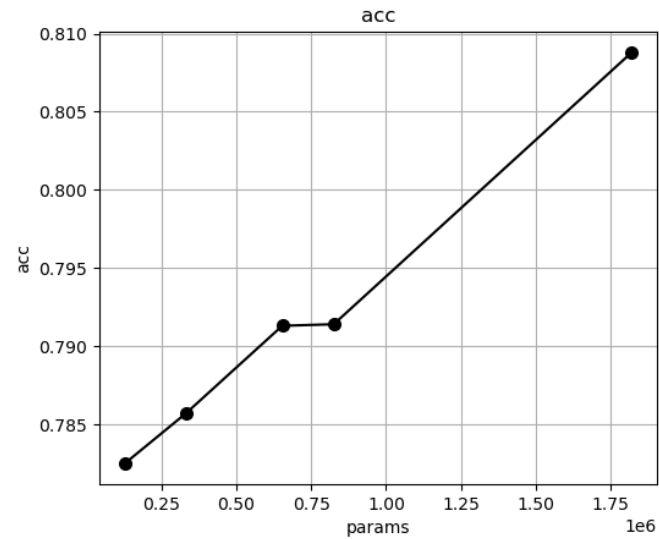
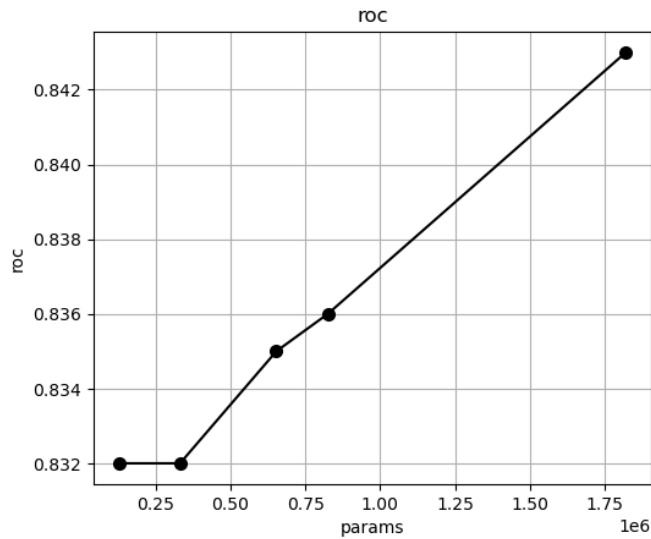


$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

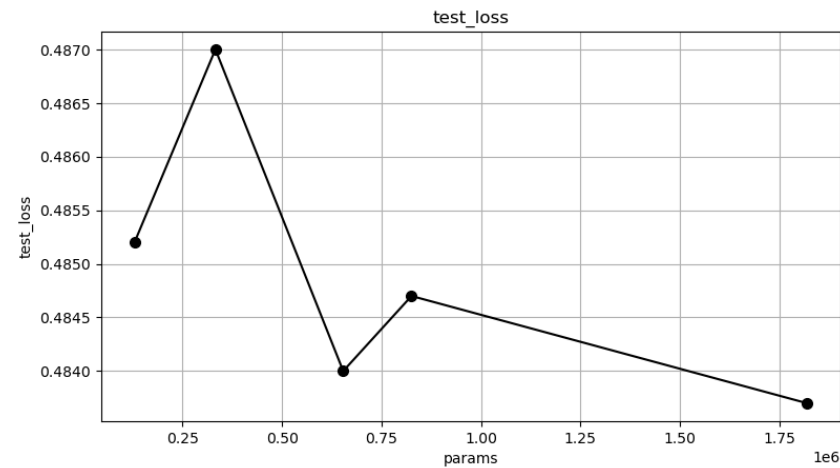
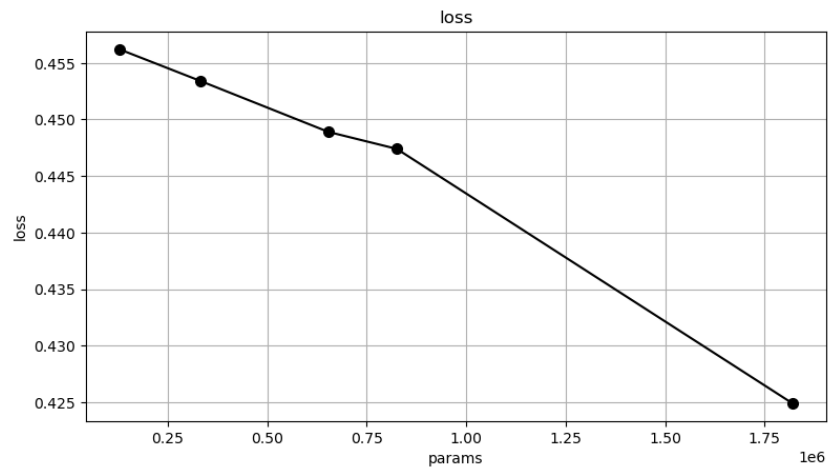
loss - функция потерь

auc - интеграл от roc-кривой

# Процесс работы. MLP(2)

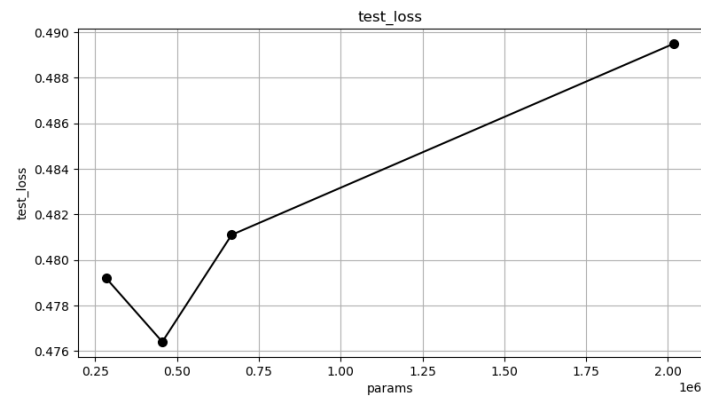
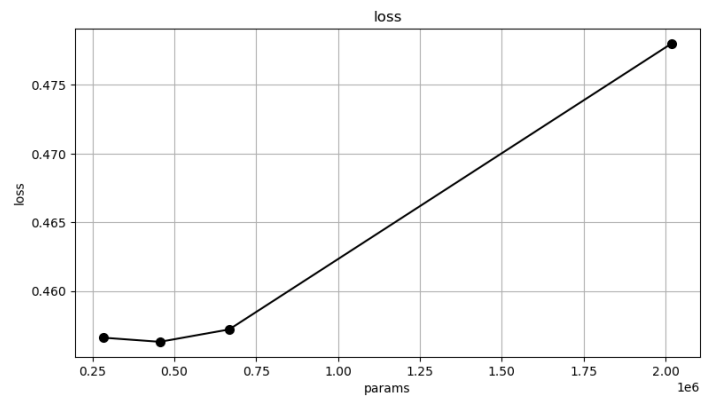
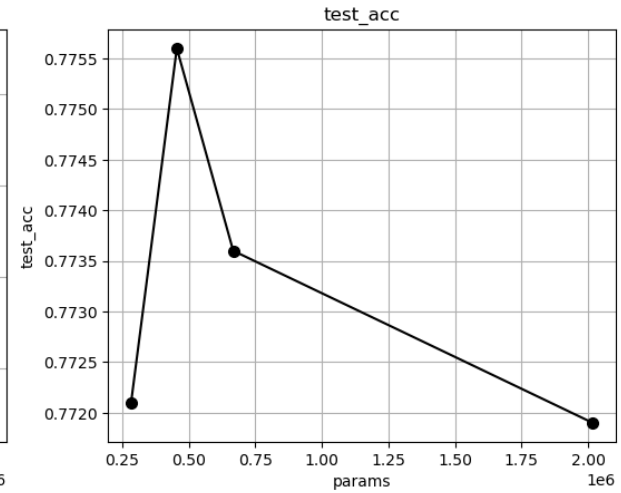
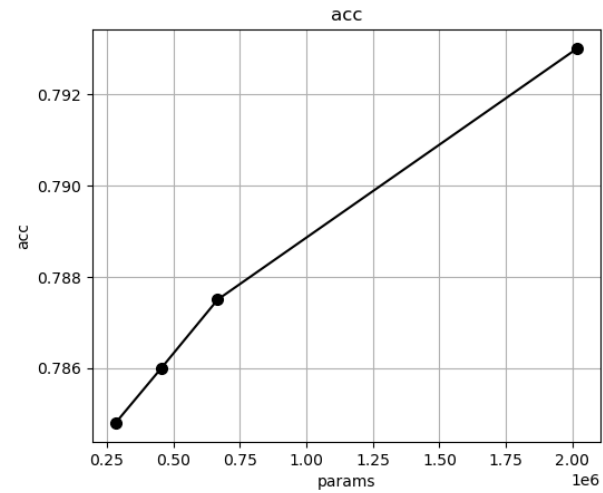
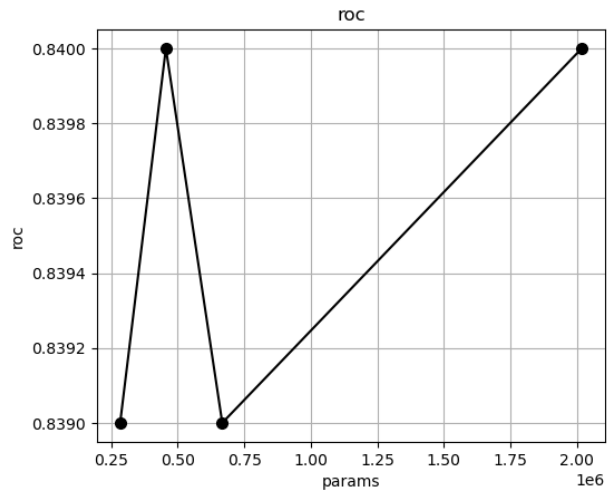


Двухслойная  
сеть



Функция активации - ReLU  
Batch = 512  
LearningRate= 0.01  
L2 - регуляризация

# Процесс работы. MLP(3)

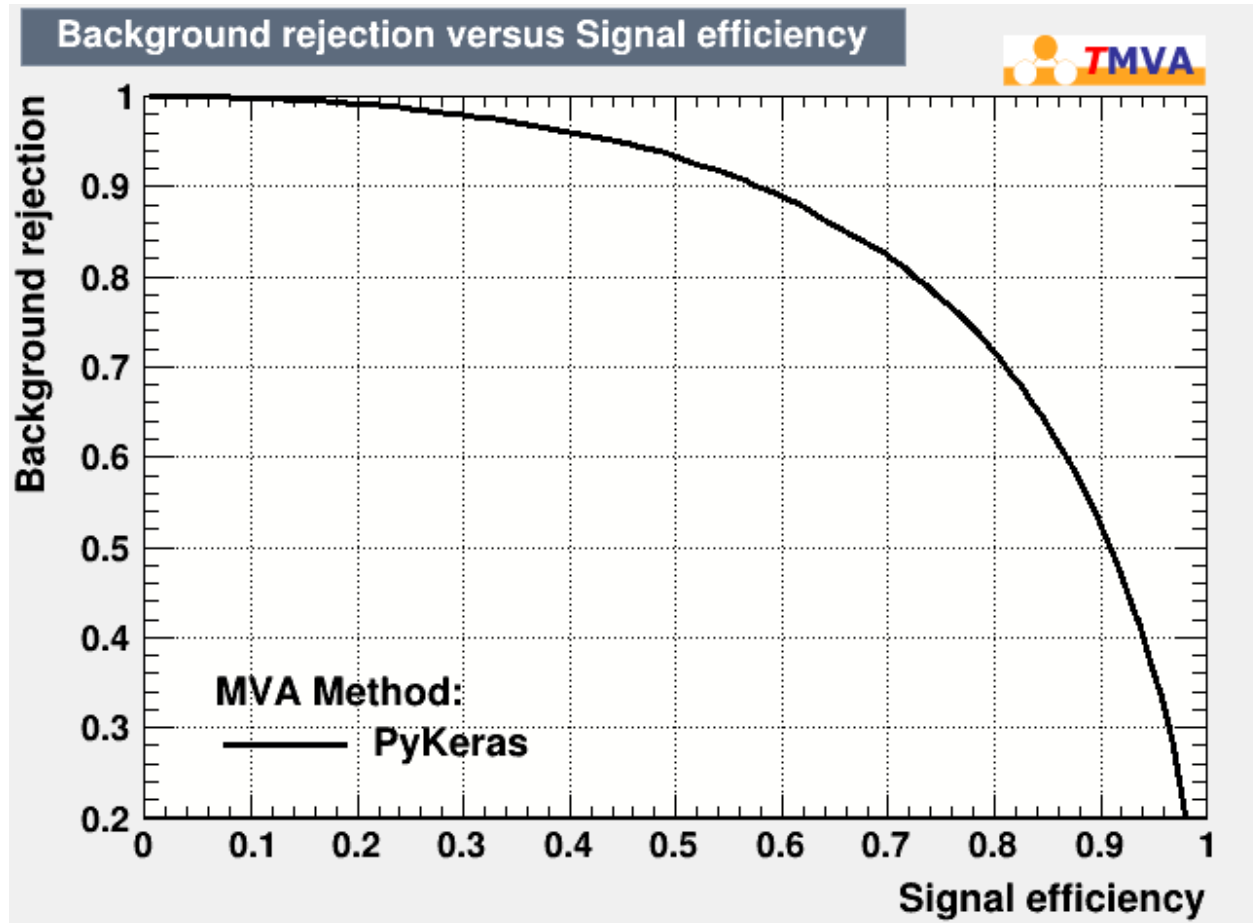


Трехслойная  
сеть

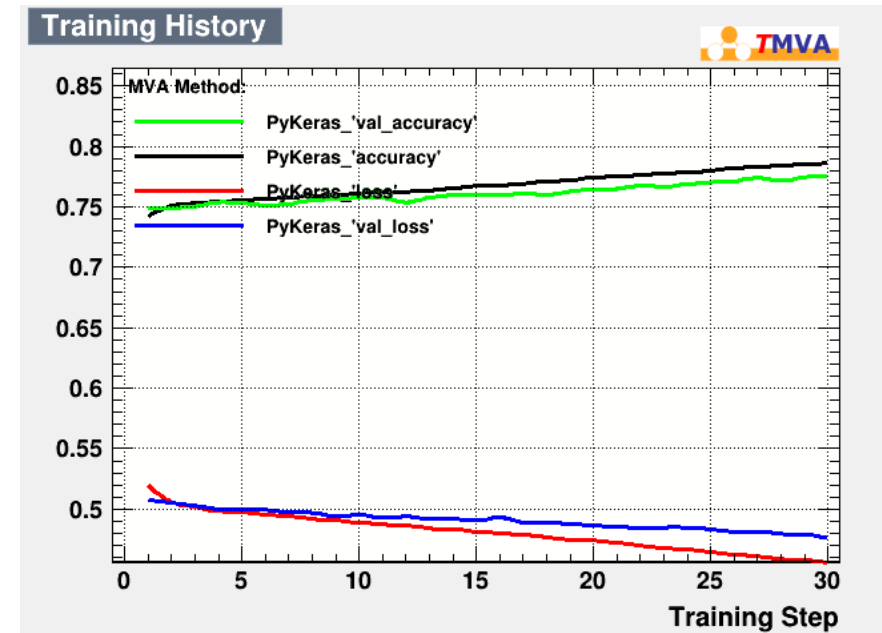
Функция активации - ReLU  
Batch = 512  
LearningRate= 0.01  
L2 + dropout



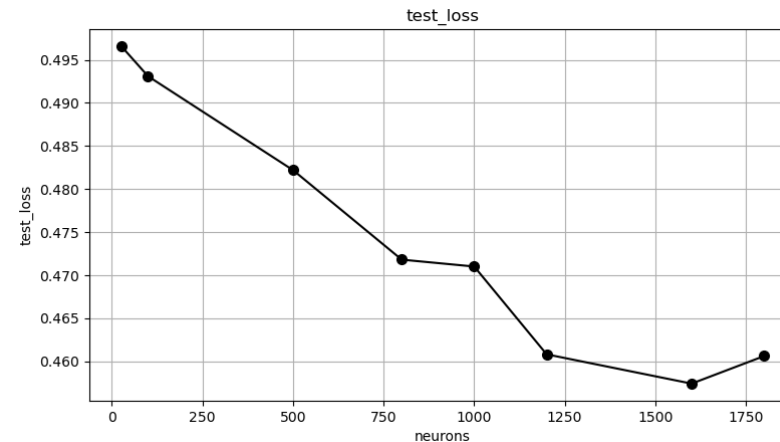
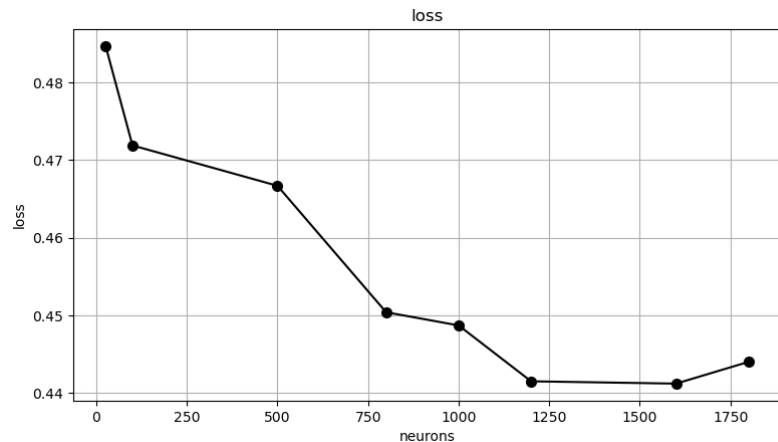
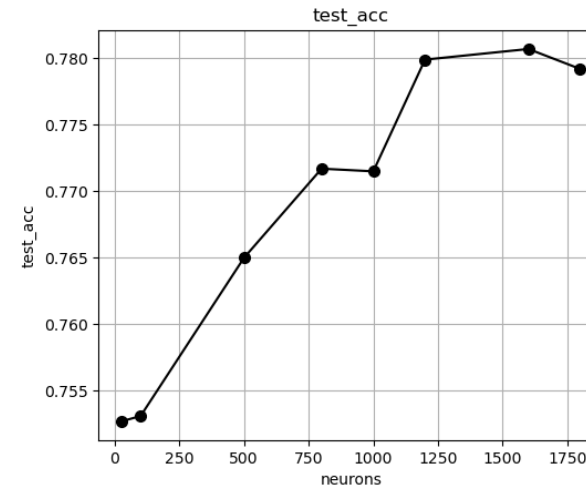
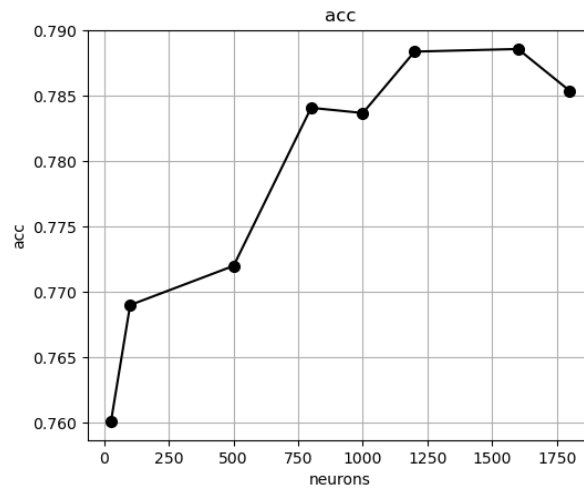
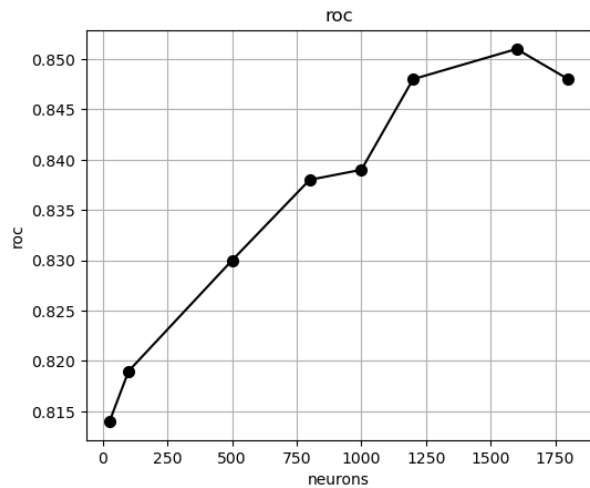
# MLP: ВЫВОД



Трехслойная сеть  
200-500-700  
AUC = 0.84



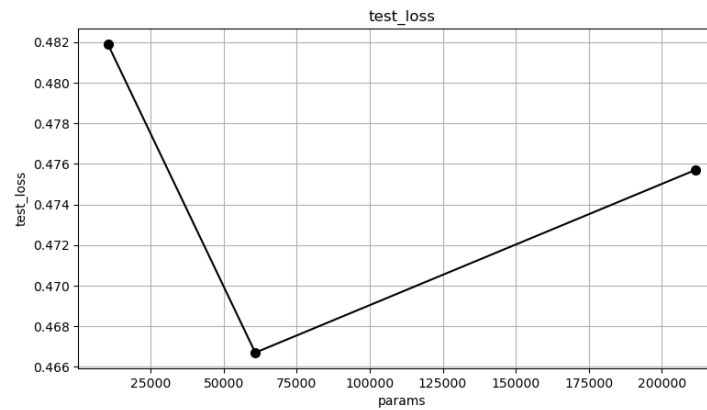
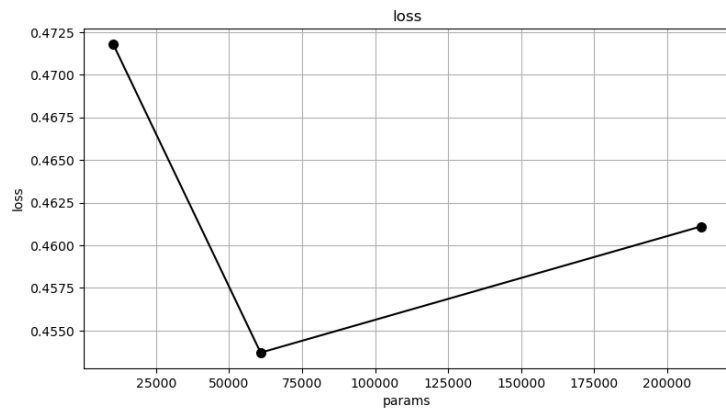
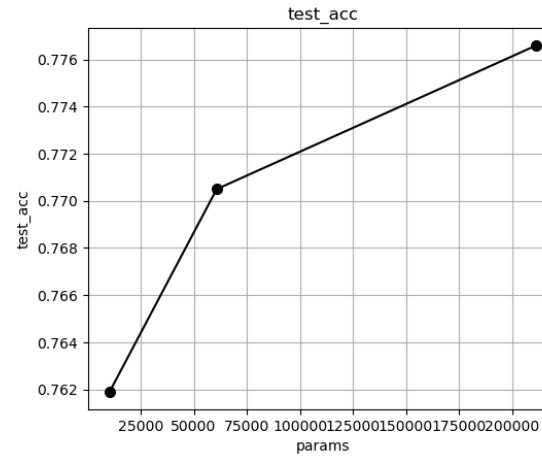
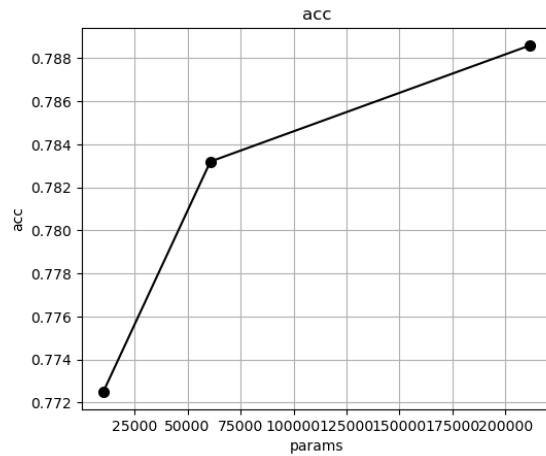
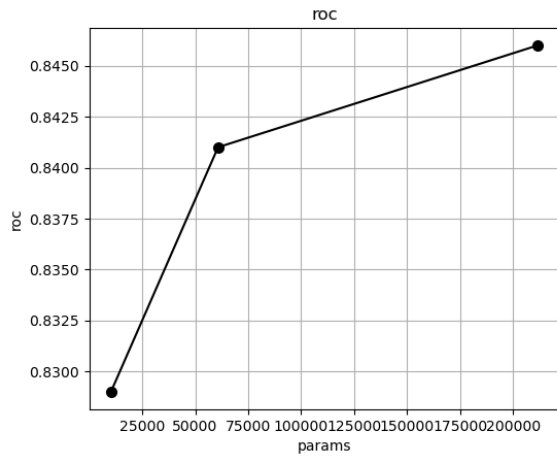
# Процесс работы. CNN(1)



**Сверточный слой  
+ Полносвязный слой**

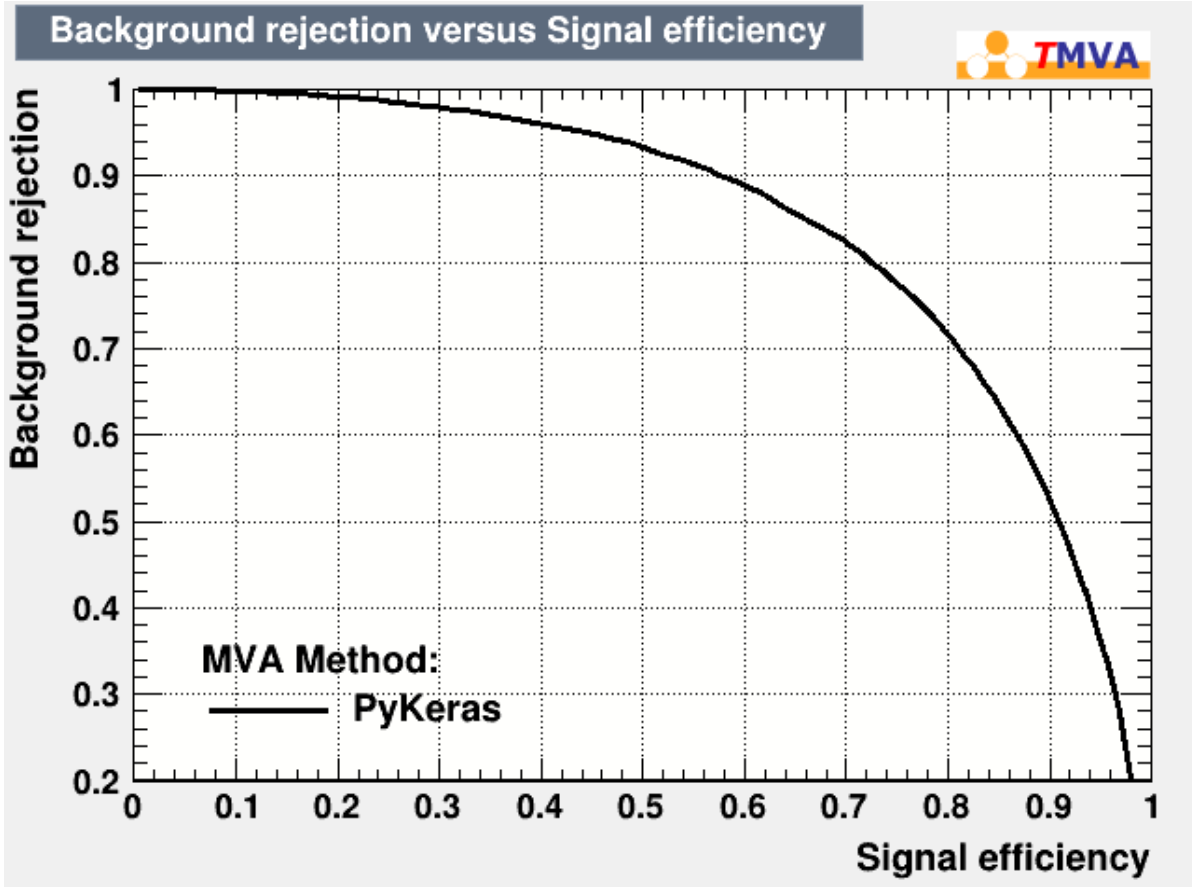
кол-во фильтров - 10  
размер фильтра - (4,1)  
размер окна - 2  
batch - 512  
функция активации - ReLU  
скорость обучения - 0.001  
+ dropout (после 200  
нейронов)

# Процесс работы. CNN(2)



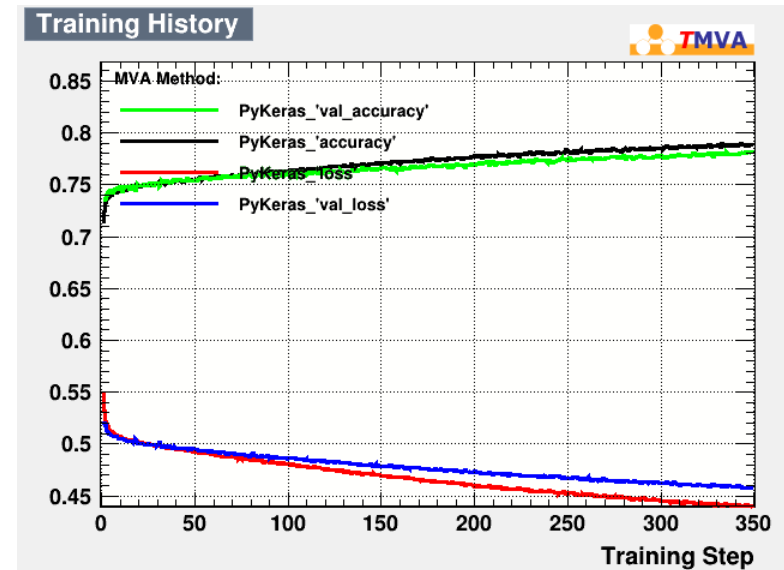
## Сверточный слой + 2 полносвязных слоя

кол-во фильтров - 10  
размер фильтра - (4,1)  
размер окна - 2  
batch - 512  
функция активации - ReLU  
скорость обучения - 0.001  
+dropout + L2



AUC = 0.850

Сверточный слой + Полносвязный слой с 1600 нейронами



# Заключение

В рамках НИР за семестр проведено ознакомление с библиотекой глубокого обучения Keras для более гибкого процесса обучения сетей, также проведено ознакомление с архитектурой многослойного перцептрона и сверточной нейронной сети для решения задачи бинарной классификации; сформированы сигнальные и фоновые деревья, содержащие выбранные дискриминирующие переменные для данной задачи. Изучены различные методы предотвращения процесса переобучения. Проведено обучение и тестирование MLP и CNN для разных наборов гиперпараметров и архитектур.

Максимальное значение AUC в прошлом семестре составляло 0.820 у модели BDT с градиентным бустингом. В данной работе получилось улучшить данную оценочную метрику, теперь максимальное значение AUC = 0.850 достигается в модели CNN с одним сверточным и одним полносвязным слоями.

В дальнейшей работе будут использоваться результаты модели, которая дала наилучшую оценочную метрику, с помощью данной модели будут классифицироваться события из реальных данных. Задача на будущее будет в сравнении эффективностей классификации данных с помощью выбранной модели и с помощью используемой ранее в анализе дискриминирующей переменной `fat_jet0_wtag50`.

**Спасибо за внимание!**

