

Глубокое обучение для геномики с помощью **DragoNN**

Участники:

Дмитрий Прокопов
Дана Зотикова
Алина Михайлова
Мария Истомина
Ксения Вальчук
Анастасия Созыкина



Руководитель:
Дмитрий Фишман

Сложность: 15 из 10

Цель проекта

Использование глубокого обучения в программном продукте **DragoNN** позволяет предсказать различные участки связывания на последовательностях генома из открытых баз данных.

Science meaning

Идентификация мест связывания транскрипционных факторов позволяет получать информацию о том, как связывание TF контролирует экспрессию генов и в конечном итоге влияет на фенотип.

Что же мы сделали?



изучили **How to train your DragoNN tutorial**



проверили работу **DragoNN** на симулированных данных



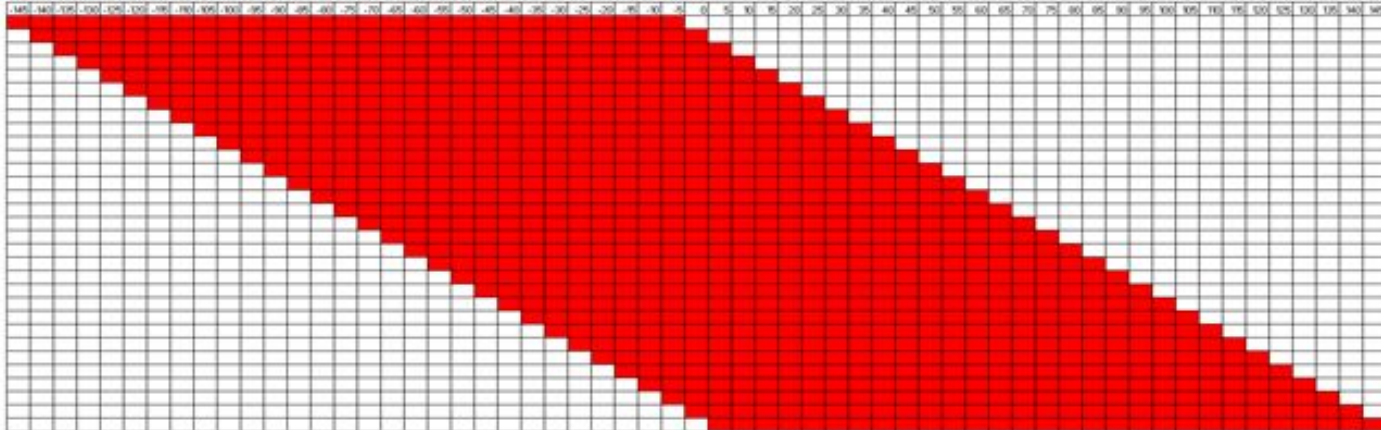
обучили модель **DragoNN** на данных Ernst et al.⁽¹⁾

Data

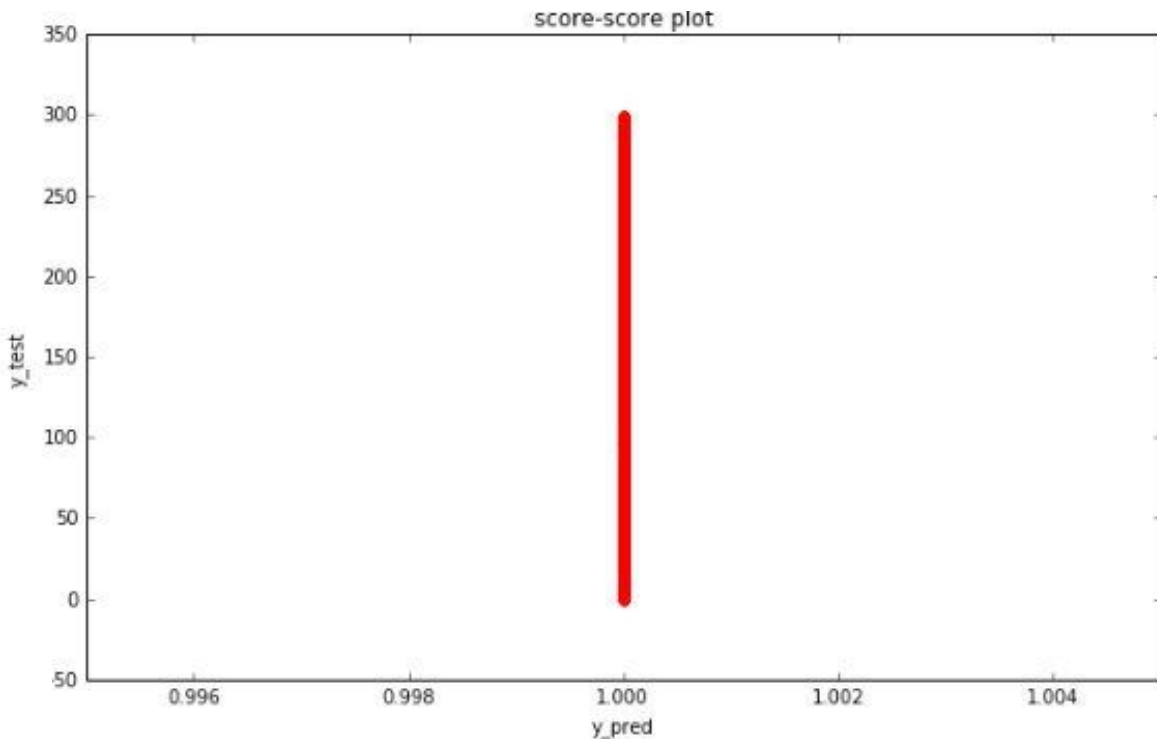
ВХОД: сиквенсы, длиной 295 bp;

ВЫХОД: counts, полученные для 2 повторных измерений (cell-type and promoter-type).

- последовательность содержит one-hot-encoded вектор длиной 145 bp, смещаемый на 5 bp для получения матрицы
- counts представляет собой количество RNA-ридов



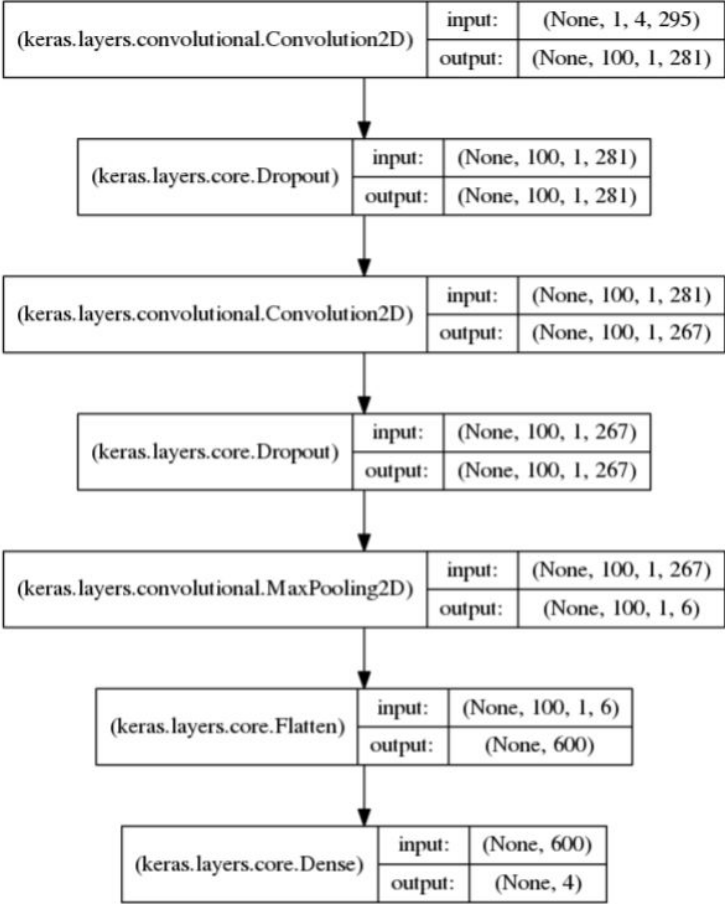
С какими трудностями мы столкнулись?



Модель выдавала константные предсказания

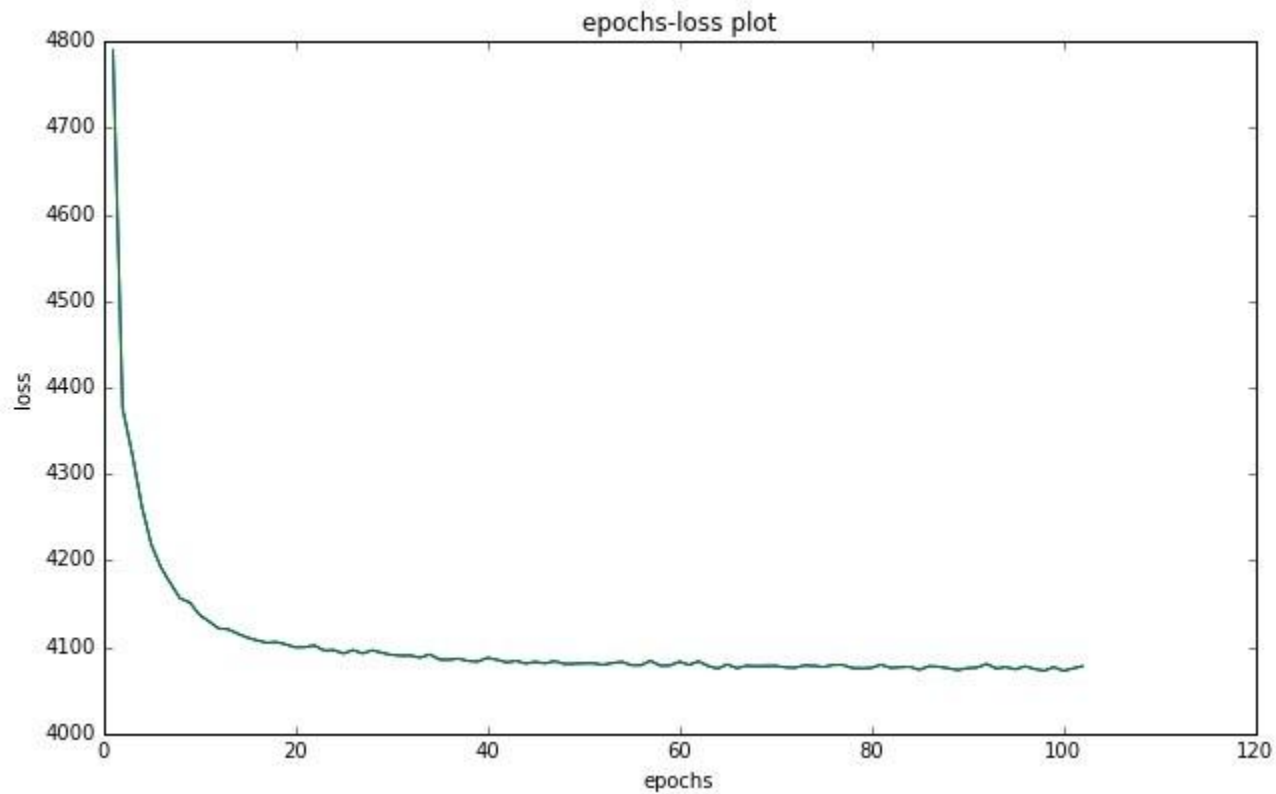


Модель

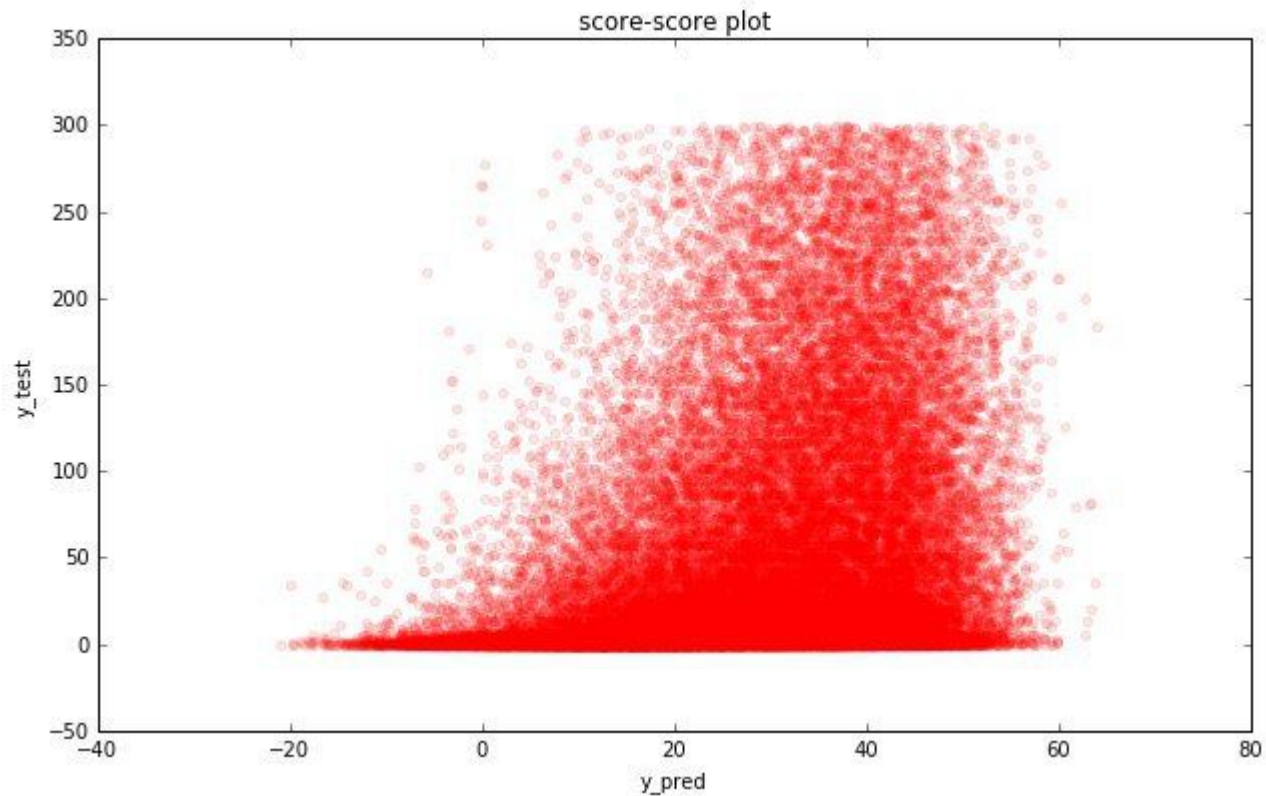


```
filter_dragonn_parameters = {  
    'seq_length': 145,  
    'num_filters': [15],  
    'conv_width': [15],  
    'dropout': 0.5,  
    'pool_width': 35  
}
```

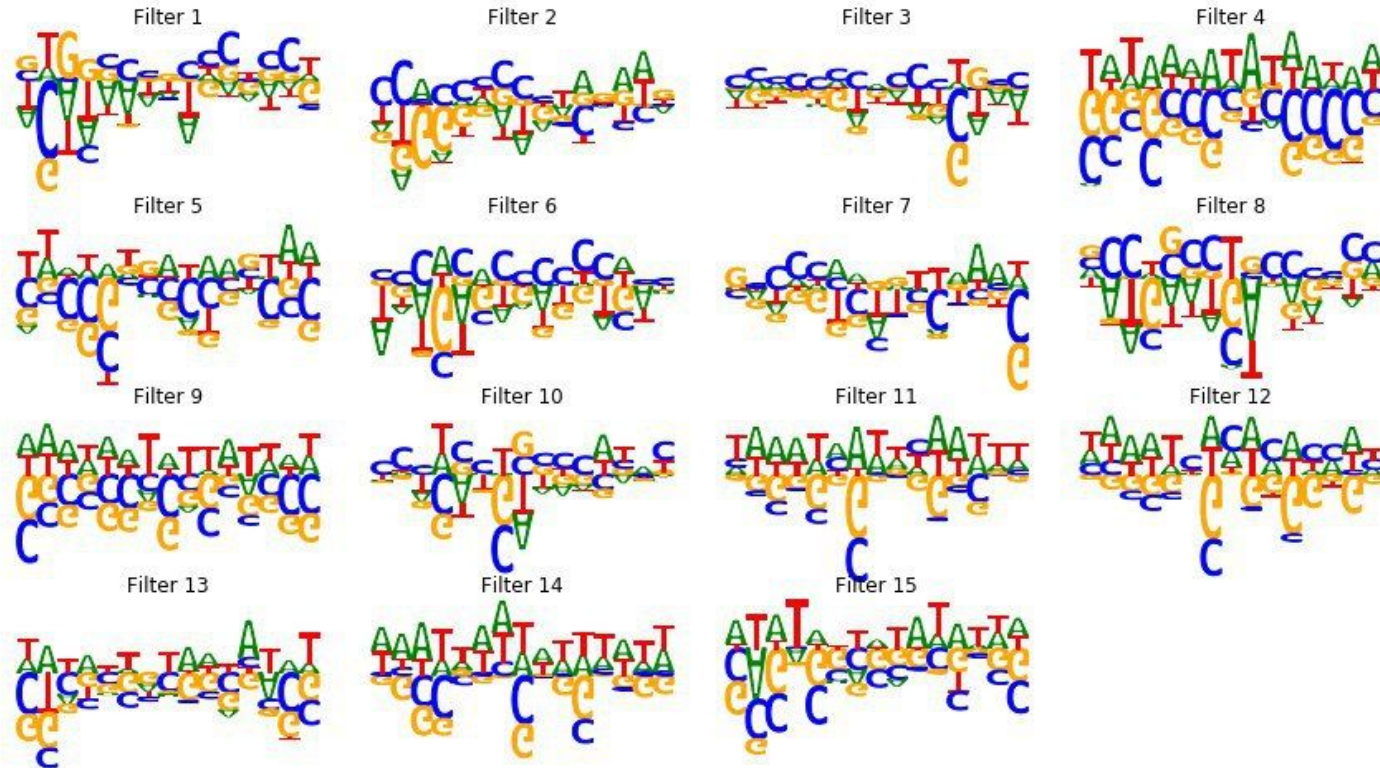
Обучение модели



Результат



Планы на будущее



Фильтры не поддаются интерпретации, основанной только на визуализации и, чтобы более эффективно изучать мотивы, в дальнейшем можно будет использовать ISM и DeepLIFT

Выводы

- мы познакомились с фреймворком DragoNN и обучили модель на данных из статьи Ernst et al., 2016
- получили неинтерпретируемую модель
- в дальнейшем следует использовать ISM и DeepLIFT

