Задачи к семинарам по курсу "Методы машинного обучения для обработки данных" кафедры физики частиц и космологии физического факультета МГУ.

Задание 7 по теме "Переобучение нейронных сетей и методы регуляризации."

7.1 Нейронная сеть с двумя промежуточными слоями. Реализовать алгоритм обучения нейронной сети с двумя промежуточными слоями для оценки длительности гамма-всплеска $T90$ по данным каталога FERMIGBRST в условиях задачи 4.2.

а) построить кривую обучения в зависимости от числа нейронов в каждом из скрытых слоев

б) сравнить точность и скорость обучения для сырых, нормированных и декоррелированных данных;

в) какой будет точность определения T90 вблизи длительности 2 секунды?

г) для какого размера сети можно увидеть эффект переобучения?

д) применить регуляризацию и проверить присутствие эффекта переобучения.

7.2 Классификация гамма-всплесков.

а) Используя методы задачи 7.1, построить классификатор гамма-всплесков, не использующий длительность в числе входных параметров. Для тренировки классификатора считать короткими всплески короче 1 секунды, а длинными – дольше 10 секунд.

б) Насколько уверенно выполняется классификация всплесков длительностью от 1 до 10 секунд? Есть ли основания предполагать существование третьего класса гамма-всплексов с другой физической природой?

7.3 Поиск корреляций между параметрами Planck. Используя алгоритм из задачи 1 выполнить оценку одного из космологических параметров Planck на основе остальных, используя данные марковских цепочек Planck, описанные в условии задачи 5.1.

а) сравнить точность и скорость обучения для сырых, нормированных и декоррелированных данных;

б) построить кривую обучения в зависимости от размера батча, используемого для обучения нейронной сети;
Замечание: кривую в этом случае строить как функцию процессорного времени, а не числа эпох.

7.4 Модификация капельной модели ядра. Фреймворки для машинного обучения позволяют находить не только оптимальные веса многослойных нейросетей, но и коэффициенты произвольных сложных функций. Вам предлагается поэкспериментировать с этим и внести собственные модификации капельной модели ядра, усложнив модель на свой вкус. Файл с данными энергии связи можно взять по ссылке: `http://ppc.inr.ac.ru/grisha_data/A_Z_deltaE.dat`. Возможные идеи:

а) Добавить ещё одно слагаемое $a_6 F(A, Z)$, где функцию F вы придумали самостоятельно, а множитель a_6 ищется по эмпирическим данным.

б) Сделать показатель степени одного из слагаемых произвольным параметром модели. Фреймворк машинного обучения должен самостоятельно вычислить градиент и найти оптимальный показатель степени в процессе обучения модели. По возможности придумайте физическое объяснение, почему показатель степени отличается от аналогичного в формуле Вайцзеккера.

в) Построить многослойную нейросеть, принимающую на вход слагаемые капельной модели ядра.

г) Использовать символическую регрессию для того, чтобы определить аналитический вид возможных поправок (пакет, PySR).

д) Использовать для оценки энергии связи сети Колмогорова-Арнольда вместо многослойного персептрона, `https://arxiv.org/abs/2404.19756`