

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

Задание 4 по теме “Скорость обучения.”

4.1 Цепочка вселенных Planck. Одним из доступных продуктов коллаборации Planck является длинная марковская цепочка наборов космологических параметров. Файл `COM_CosmoParams_fullGrid_R3.00.tgz` (9 Гб) доступен по адресу
https://wiki.cosmos.esa.int/planck-legacy-archive/index.php/Cosmological_Parameters#Parameter_Chains

Цепочка построена методом Монте-Карло марковских цепей, при котором частота пребывания цепи в каждой точке пространства параметров оказывается пропорциональной функции правдоподобия, характеризующей согласие модели Вселенной с заданными параметрами с наблюдениями Planck. В каждой строке текстового файла последовательно записаны: вес, функция правдоподобия и космологические параметры. Вес соответствует числу шагов, в течение которого цепочка находилась в данном состоянии.

- а) вычислить матрицу ковариации для космологических параметров;
- б) используя методы, описанные в заданиях к лекциям 2,3, построить регрессию для оценки функции правдоподобия, исходя из значений космологических параметров;
- в) исследовать зависимость точности алгоритма и скорости обучения от способа подготовки данных;
- г) исследовать зависимость точности алгоритма и скорости обучения от функции стоимости и функции активации;