Оценка за задание: для пересдачи максимальный балл за задание — 8 баллов. 8 баллов соответсвует тому, что все требования задания выполнены, методы находят решения с высокой точностью ($1e{-}16$) быстро (на датасетах, сравнимых по размеру с a1a время работы не должно для всех методов должны быть меньше 2-3 минут).

В этом задании вам предстоит научиться решать Poisson Regression разными методами, рассмотренными на протяжении курса. Необходимо реализовать скрипт, умеющий находить решение на произвольном датасете с помощью следующих методов:

1. Gradient descent
2. Newton method (в реализации нужно учесть ситуации с плохой обусловленностью и уметь обрабатывать наличие неположительных собственных чисел в гессиане)
3. Hessian-Free Newton
4. BFGS
5. LBFGS
6. ADAM
7. Proximal или любой другой "быстро"-сходящийся метод для решения с l_1-регуляризацией

Минимальные требования к скрипту:

1. Запускается и работает без вмешательства с моей стороны (если у меня скрипт падает по любой причине без внятной диагностики, дальнейшая проверка задания не гарантируется)
2. Выдает значение функции в оптимальной точке
3. Выдает оптимальную точку в виде массива
4. Норму градиента в точке оптимума
5. Значение критерия остановки в точке оптимума
6. Для разреженных решений — число ненулевых компонент
7. Работать с нулевыми параметрами регуляризации
8. Корректно решать поставленную в условии задачу оптимизации (т.е. если вы где-то делаете нормировку признаков, например, то должны правильно перенормировать и параметр регуляризации)

Решения необходимо реализовать на Python 3.6+, используя scipy и numpy.

Интерфейс скрипт: аналогично домашним заданиям (line-search стратегию можете выбрать самостоятельно одну) Данные находятся в libsvm-формате, их можно загружать в питон с помощью sklearn.datasets.load_svmlight_file

Модель пуасоновской регрессии

Дано N пар (x_i, y_i), где $x \in \mathbb{R}^n$, $y_i \in \mathbb{N}_{\geq 0}$. x принятно называть наблюдениями (примерами, etc), y_i лейблами/метками/etc. В случаи пуасоновской регрессии предполагается, что y_i это случайные величины, имеющиеся Пуасоновское распределение, со средним, равным $w^T x_i$.

Рассматривается следующая модель зависимости y от x:

$$p(y|x) = \frac{\lambda(x)^y}{e}^{\lambda} y!, \tag{1}$$

где

$$\lambda_w(x) = e^{w^T x} \tag{2}$$

Для нахождения параметра w предлагается решать следующую оптимизационную задачу:

$$w = \arg\max_{w \in \mathbb{R}^n} \frac{1}{N} \sum_{i=1}^{N} (y_i w^T x_i - e^{w^T x_i}) + \frac{\lambda_0}{2} ||w||_2^2 + \lambda_1 ||w||_1 \tag{3}$$

Замечание: в описании модели я предполагаю, что в x последняя координата тождественна равна 1. Первые $n-1$ компонент вектора это признаки, с помощью которых описывается зависимость, последняя для удобного задания константной поправки. Для тех, кто не знаком с таким способом записи — представьте обычную линейную зависимость $\alpha x_0 + \beta$. Здесь 2 параметра, их надо искать. Для записи в векторном виде удобно добавить dummy-переменную x_1, всегда равную 1

Важное замечание: в данном задании лейбы принимают любое натуральное число, а не являются бинарными

Тестовые данные: в этом задании датасет для тестов для вас будет не доступен, так что необходимо будет подготовить свой набор данных для тестов и проверок того, что ваш код работает. При проверки будет использоваться набор датасетов, подготовленных мной. Код должен будет работать на моих данных, а не на ваших тестовых примерах