Задание №1.2. Выпуклость обобщенных линейных моделей

В этом упражнении мы исследуем некоторые полезные свойства обобщенных линейных моделей, в частности те, которые следуют из использования ими распределений из экспоненциального семейства для моделирования выходных данных. Чаще всего ОЛМ обучаются с использованием отрицательного логарифмического правдоподобия (ОЛП[1]) в качестве функции потерь (будем далее называть ее функцией ОЛП потерь). С математической точки зрения это эквивалентно оценке максимального правдоподобия (т.е. максимизация логарифмического правдоподобия эквивалентна минимизации отрицательного логарифмического правдоподобия). Наша цель состоит в том, чтобы показать, что функция ОЛП потерь для ОЛМ является выпуклой функцией от параметров модели.

Для нас это важно, потому что выпуклая функция – это функция, для которой любой локальный минимум также является глобальным минимумом, и существует большой и хорошо разработанный математический аппарат, позволяющий эффективно искать оптимум для разных типов выпуклых функций с помощью различных алгоритмов, таких как, например, стохастический градиентный спуск.

Напомним, что экспоненциальное семейство распределений имеет следующую формулу (плотности) вероятности:

$$p(y;\eta) = b(y)\exp\big(\eta^T T(y) - a(\eta)\big),$$

где η – это естественный параметр распределения. Более того, в ОЛМ η моделируется как $\theta^T x$, где $x \in \mathbb{R}^n$ – вектор признаков объекта, а $\theta \in \mathbb{R}^n$ – обучаемые параметры модели. Разобьем весь процесс на шаги и выполним их один за другим. Наш план состоит в том, чтобы показать, что вторая производная (т.е. гессиан) функции ОЛП-потерь по параметрам модели является положительной полуопределенной матрицей для всех значений параметров модели. По пути мы также выведем некоторые полезные свойства экспоненциального семейства распределений.

Предположим, что $p(Y|X;\theta) \sim ExponentialFamily(\eta)$, где $\eta \in \mathbb{R}$, а $T(y) = y$. Для удобства ограничимся случаем, когда η – скаляр. Тогда мы можем упростить формулу экспоненциального семейства:

$$p(y;\eta) = b(y)\exp\big(\eta y - a(\eta)\big).$$

Вопрос №1

Выведите формулу математического ожидания распределения. Покажите, что $E[Y;\eta] = \frac{\partial}{\partial x} a(\eta)$ (заметьте, что $E[Y;\eta] = E[Y|X;\theta]$, так как $\eta = \theta^T x$). Другими словами, покажите, что ожидаемое значение распределения из экспоненциального семейства есть первая производная нормировочного коэффициента, взятая по естественному параметру.

Подсказка: начните с того факта, что $\frac{\partial}{\partial \eta}\int p(y;\eta)dy = \int \frac{\partial}{\partial \eta} p(y;\eta)dy$.

Вопрос №2

Далее выведите формулу дисперсии распределения. В частности, покажите, что $Var(Y;\eta) = \frac{\partial^2}{\partial \eta^2} a(\eta)$ (опять, заметьте, что $\mathrm{Var}(Y;\eta) = Var(Y|X;\theta)$). Другими словами, покажите, что дисперсия распределения из экспоненциального семейства есть вторая производная нормировочного коэффициента по естественному параметру.

---

[1] NLL, negative log-likelihood

Подсказка: попробуйте использовать результаты, полученные в процессе решения предыдущего вопроса.

Вопрос №3

Наконец, выпишите функцию потерь $\ell(\theta)$ – отрицательное логарифмическое правдоподобие распределения, как функцию от θ. Затем, посчитайте гессиан этой функций по параметру θ и покажите, что он всегда является положительным полуопределенным. Это завершит доказательство того, что функция ОЛП потерь для ОЛМ является выпуклой.

Подсказка 1: используйте правило суперпозиции и все полученные ранее результаты, чтобы упростить вывод.

Подсказка 2: вспомните, что дисперсия любого распределения неотрицательна.

Основные выводы, которые мы можем получить, решив это упражнение:

- Любая обобщенная линейная модель выпукла по своим модельным параметрам.
- Экспоненциальное семейство распределений вероятностей с математической точки зрения является удобным. В то время как вычисление ожидаемого значения и дисперсии для произвольных распределений в общем случае подразумевает интегрирование, удивительным образом для экспоненциального семейства мы можем вычислить их, используя дифференцирование, которое, как правило, легче интегрирования.