

Машинное обучение, ФКН ВШЭ

Теоретическое домашнее задание №11

Задача 1. Пусть целевая переменная имеет отрицательное биномиальное распределение с фиксированным параметром r :

$$p(y | \theta(x)) = C_{y+r-1}^y \theta(x)^y (1 - \theta(x))^{r-y}.$$

Запишите оптимизационную задачу поиска вектора весов модели w для соответствующей обобщенной линейной модели (для метода максимального правдоподобия), выразите значение параметра $\theta(x)$ через оптимальное значение w^* и найдите матожидание ответов y , обусловленное натуральным параметром $\eta(x)$ через дифференцирование логарифма нормировочной константы экспоненциальной формы распределения. Совпадает ли оно с $\theta(x)$ и почему? Как будет выглядеть итоговый алгоритм прогнозирования $a(x)$?

Задача 2. Пусть целевая переменная имеет распределение Парето с фиксированным параметром k :

$$p(y | \alpha(x)) = \frac{\alpha(x) k^{\alpha(x)}}{y^{\alpha(x)+1}}.$$

Запишите оптимизационную задачу поиска вектора весов модели w для соответствующей обобщенной линейной модели (для метода максимального правдоподобия), выразите значение параметра $\alpha(x)$ через оптимальное значение w^* и найдите матожидание достаточной статистики $s(y)$, обусловленное натуральным параметром $\eta(x)$ через дифференцирование логарифма нормировочной константы. Совпадает ли оно с $\alpha(x)$ и почему? Какой из вариантов является лучшим кандидатом на роль алгоритма $a(x)$?

Подсказка. Напомним, что в записи экспоненциальной формы распределения может фигурировать не y , а некоторая статистика $s(y)$.