

Самостоятельная работа

Предметная область

Предположим, что мы хотим предсказывать количество звонков в техподдержку. Основные гипотезы:

- количество звонков линейно зависит от множества параметров (время, количество сбоя в работе серверов за последнее время, наличие рекламных акций и т.п.)
- распределено по закону Пуассона.

$$p(y; \lambda) = \frac{e^{-\lambda} \lambda^y}{y!}$$

Также, вводим условие, что модель должна дообучаться в реальном времени, то есть после каждого нового звонка. Требуется написать алгоритм обучения и предсказания новых значений.

Подсказки

- Покажите, что распределение Пуассона из экспоненциального семейства

$$p(y; \eta) = b(y) \exp \{ \eta^T T(y) - a(\eta) \}$$

и выпишите чему равны $b(y), \eta, T(y), a()$

- Выпишите вид гипотезы $h_\theta(x)$ для данной модели.
- Постройте функцию логарифма максимального правдоподобия

$$\log \prod_{i=1}^m p(y^{(i)} | x^{(i)}; \theta)$$

- Используя правило для стохастического градиентного подъема

$$\theta_j := \theta_j + \alpha \frac{\partial J(\theta)}{\partial \theta_j},$$

выпишите алгоритм для обновления θ