

Pareto/NBD в «sklearn’овских» терминах

Математическая модель (ММ)	Линейная регрессия $Y = k * x + bias$	Pareto/NBD
Параметры ММ модели, которые нужно найти, для того, чтобы $f(x) = Y\_pred,$ причем $(Y\_pred - Y\_true) \rightarrow 0$	k, bias	Гиперпараметры?: $\lambda$ – transaction rate (отвечает за частоту транзакций, не зависит от клиента) $\mu$ – churn rate (отвечает за время жизни клиента, не зависит от клиента)  Параметры: Параметры гамма распределения, описывающие различия между клиентами в их поведении относительно совершения транзакций $r$ – shape parameter $\alpha$ - scale parameter Параметры гамма распределения, описывающие различия между клиентами в показателях отсева (оттока) $s$ – shape parameter $\beta$ - scale parameter
Признаки	Матрица «объект – признак»	Матрица «клиент – признак» Для описания клиента достаточно знать: 1. Число транзакций в период времени от начала наблюдения до текущего момента 2. Время последней покупки, между 1ой и последней покупкой. Т.е. нам надо знать всего 2 информации – давность и частота покупок
Целевая переменная, что прогнозируем, выход модели	Y (какое то число)	$P(X)$ – вероятность наблюдения x транзакций в период времени t в будущем $E(X)$ - ожидаемое число транзакций в прогнозный период для клиента с наблюдаемым поведением
Функционал ошибки	минимизируем $\Phi = (y\_true - y\_pred)^2$	максимизируем Функцию максимального правдоподобия $\frac{\lambda^x \mu}{\lambda + \mu} e^{-(\lambda + \mu) t_x} + \frac{\lambda^{x+1}}{\lambda + \mu} e^{-(\lambda + \mu) T}$
Метод оптимизации	Градиентный спуск, SGD, Adam, adagrad и др.	Hamiltonian Monte Carlo (НМС) (в stan) или др.алгоритмы численной оптимизации *

\*Функция правдоподобия, связанная с моделью Парето / NBD, довольно сложна и требует многочисленных вычислений гипергеометрической функции Гаусса. Множественные оценки гауссовской гипергеометрии очень требовательны с вычислительной точки зрения. Кроме того, точность некоторых *числовых процедур*, используемых для оценки этой функции, может существенно варьироваться в пространстве параметров; Это может вызвать серьезные проблемы для методов численной оптимизации, поскольку они ищут максимум функции правдоподобия.