Решение прикладных математических задач в продуктовой аналитике

Маргарян Ашот

27 июня 2025 г.

Введение

В данном разделе собраны примеры задач, в которых применяются углублённые методы математической статистики, функционального анализа и теории вероятностей для решения прикладных задач в сфере продуктовой аналитики.

1 Bootstrap-интервалы при малых выборках

Условие

Тестируется новый баннер в мобильном банке. Выборка малая: 80 пользователей, из них 12 кликнули на баннер.

Цель

Оценить доверительный интервал конверсии без использования нормального приближения.

Решение

Пусть $X_i = 1$, если пользователь кликнул, $X_i = 0$ — иначе. Проведём $B = 10000$ бутстрэп-выборок с возвращением и посчитаем распределение среднего:

$$\hat{p}^{(b)} = \frac{1}{n} \sum_{i=1}^{n} X_i^{*(b)}$$

где $X^{*(b)}$ — b-я бутстрэп-выборка.

Точечная оценка:

$$\hat{p} = \frac{12}{80} = 0.15$$

После генерации бутстрэп-распределения определим 2.5% и 97.5% квантили:

$$CI_{95\%}^{\text{bootstrap}} = [0.081, 0.225]$$

2 Мультиколлинеарность и регуляризация

Условие

В модели кредитного скоринга признаки "доход "сумма кредита "срок"сильно коррелированы.

Цель

Стабилизировать линейную модель с помощью регуляризации (ридж-регрессия).

Решение

Рассматриваем линейную модель:

$$\hat{y} = X\beta$$

В обычной МНК:

$$\hat{\beta}_{OLS} = (X^\top X)^{-1} X^\top y$$

Из-за мультиколлинеарности матрица $X^\top X$ почти вырождена. Решаем это через ридж-регрессию:

$$\hat{\beta}_{\text{ridge}} = (X^\top X + \lambda I)^{-1} X^\top y$$

Выбор параметра λ — по кросс-валидации. Эмпирически:

$$\lambda^* = 0.12 \Rightarrow \text{улучшение RMSE на } 15\%$$

3 Оптимизация удержания через функциональную норму

Условие

Хотим смоделировать churn-функцию $f(t)$ — вероятность оттока клиента в момент времени t (неделя), и подобрать оптимальную стратегию удержания.

Цель

Минимизировать функционал, зависящий от $f(t)$ и стоимости коммуникаций.

Решение

Рассматриваем функционал:

$$J[f] = \int_0^T \left(c(t) f(t) + \lambda \left(\frac{df}{dt} \right)^2 \right) dt$$

где $c(t)$ — стоимость удержания в момент t, λ — параметр сглаживания.

Ищем минимум функционала в пространстве $H^1([0,T])$ (собственный функциональный анализ).

Решение методом Эйлера–Лагранжа:

$$\frac{d}{dt}\left(2\lambda\frac{df}{dt}\right) = c(t) \Rightarrow \frac{d^2 f}{dt^2} = \frac{c(t)}{2\lambda}$$

Пример: $c(t) = 1 + 0.1t$, $\lambda = 0.5$.
Тогда:

$$f(t) = \frac{1}{2\lambda}\left(t + 0.05t^2\right) + At + B$$

Параметры A, B — из граничных условий (например, $f(0) = 1$, $f(T) = 0$). Получаем аналитическое решение оптимального распределения удержания.

4 Интерпретация скрытых состояний в HMM для моделирования поведения

Условие

Хотим сегментировать пользователей по паттерну действий в приложении. Поведение дискретно: `(просмотр, клик, покупка)`.

Цель

Смоделировать поведение через скрытую марковскую модель (HMM) и интерпретировать скрытые состояния.

Решение

Наблюдаемые состояния $O = \{V, C, P\}$, скрытые — $S = \{, , \}$.

Используем алгоритм Баума–Уэлша для оценки параметров модели: - Матрица переходов A - Матрица эмиссий B - Начальные вероятности π

После обучения получаем:

$$A = \begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix} \quad B = \begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.6 & 0.3 & 0.1 \\ 0.7 & 0.2 & 0.1 \end{bmatrix}$$

Интерпретация: состояние S_1 — это покупатели, S_3 — случайные визитёры. Используем декодирование Витерби для прогнозирования поведения нового пользователя.