

Анализ спроса на рынке соков в Перми

Димитров В.К

Новосибирский Государственный университет

27 апреля 2022 г.

1 Постановка задачи

- 1 Постановка задачи
- 2 Эконометрический подход

Содержание

- 1 Постановка задачи
- 2 Эконометрический подход
- 3 Машинное обучение

- 1 Постановка задачи
- 2 Эконометрический подход
- 3 Машинное обучение
- 4 Смешанная модель

- 1 Постановка задачи
- 2 Эконометрический подход
- 3 Машинное обучение
- 4 Смешанная модель
- 5 Вопросы

Постановка задачи

- ❶ Задача: предсказание спроса на рынке соков в Перми.
- ❷ Количество признаков 25 (после обработки 40).
- ❸ Целевая переменная - количество продаж.
- ❹ Методы:
 - ❶ Эконометрический подход
 - ❷ Инженерный подход
 - ❸ Смешанный подход

Распределение Пуассона

ξ - это случайная величина, которая имеет распределение Пуассона.
Тогда

$$P\{\xi = k\} = \frac{\exp(-\lambda) \times \lambda^k}{k!}, k = \{0, 1, \dots\}$$

Эконометрическая спецификация

Эконометрическая спецификация предложения , заключается в :

$$\lambda = \exp(x'\beta)$$

Эконометрическая спецификация

Эконометрическая спецификация предложения , заключается в :

$$\lambda = \exp(x' \beta)$$

Модель для y_i покупки

Пусть $y_i = k$, тогда наша модель перепишется в виде:

$$P(\xi = y_i) = \frac{\exp(\exp(x' * \beta)) \times \exp(x' \beta)^{y_i}}{y_i!}$$

Максимум правдоподобия

Для оценки используется оценка максимального правдоподобия. Тогда задача переписывается в виде:

$$\max_{\beta} \frac{1}{N} \sum_{i=1}^N \{-\exp(x_i' \beta) + y_i x_i' \beta - \ln(y_i)!\}$$

Но...

Задача не имеет явного решения

- 1 Поиск по сетке (Grid search)
- 2 Градиентный спуск (Gradient method)
- 3 Метод Ньютона-Рафсона (Newton–Raphson Method)

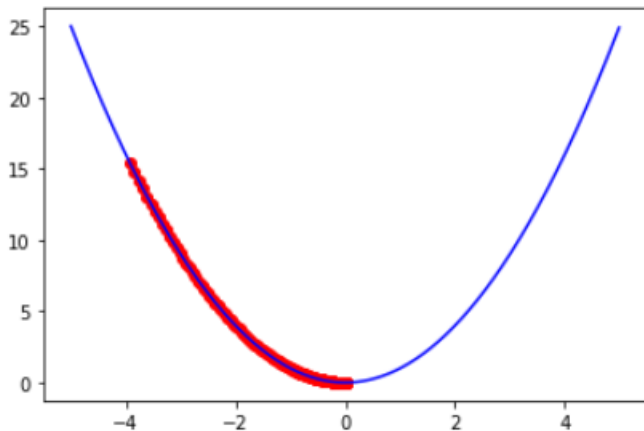
Градиентный метод

Градиентные методы (Gradient Method). Новые значения признаков пересчитываются по формуле:

$$\hat{\beta}_{i+1} = \hat{\beta}_i + A_i g_i$$

Численные методы (Пример)

Куда дошли:, $-6.7318694288638116e-09$
А минимум θ



Описание метода

Так как наша функция глобально вогнутая, тогда метод НР работает достаточно хорошо. Градиентный спуск модифицируется:

$$\hat{\beta}_{i+1} = \hat{\beta}_i - H_i^{-1} g_i$$

Описание метода

Так как наша функция глобально вогнутая, тогда метод НР работает достаточно хорошо. Градиентный спуск модифицируется:

$$\hat{\beta}_{i+1} = \hat{\beta}_i - H_i^{-1} g_i$$

$H_i = \frac{\partial^2 Q_N(\beta)}{\partial \beta \partial \beta'} \Big|_{\hat{\beta}_i}$ - гессиан функции $Q(\beta)$ в точке $\hat{\beta}$.

Проблемы

- 1 Как оптимизировать Гессиан?

Проблемы

- 1 Как оптимизировать Гессиан?
- 2 Как выбирать подвыборку?

Проблемы

- 1 Как оптимизировать Гессинан?
- 2 Как выбирать подвыборку?
- 3 Сколько итераций и инициализация?

Проблемы

- 1 Как оптимизировать Гессиан?
- 2 Как выбирать подвыборку?
- 3 Сколько итераций и инициализация?
- 4 Learning rate

Проблемы

- 1 Как оптимизировать Гессиан?
- 2 Как выбирать подвыборку?
- 3 Сколько итераций и инициализация?
- 4 Learning rate
- 5 Написать программу

Расширение спецификации

- Пуассон с раздутым нулем (zero inflated poisson)
- Мультиномиальная логистическая модель
- Спецификация для параметра λ

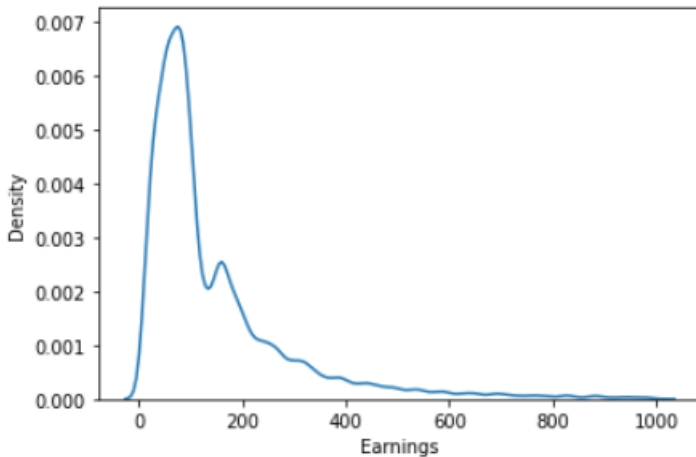
- Логистическая регрессия
- Random Forest
- GradientBoostingClassifier

Модель	Параметры	Точность(до, после)
Логистическая	lr, alpha	0.74, 0.76
Random Forest	Много	0.72, 0.77
GBR	Много	0.84

Мой notebook

- Можно взять предсказание эконометрической модели и подать на вход любой модели ML (stacking).
- Можно использовать методы кластеризации, для выявления новых признаков для эконометрических моделей.

Пример кластеризации



Научный блок

- Эластичность
- Интерпретация коэффициентов
- Индексы
- Получение состоятельных оценок

Научный блок

- Эластичность
- Интерпретация коэффициентов
- Индексы
- Получение состоятельных оценок

Бизнес блок

- Как перевести метрику качества в метрику бизнеса ?
- Что может быть интересно бизнесу?

Готов ответить на ваши вопросы.

Спасибо за внимание!