Министерство науки и высшего образования Российской Федерации

Южно-Уральский государственный университет

(национальный исследовательский университет)

Высшая школа электроники и компьютерных наук

Кафедра «Информационно-аналитическое обеспечение управления в социальных и экономических системах»

ОТЧЕТ  
по заданию №4  
по дисциплине «Технологии и системы интеллектуально-аналитической обработки данных в экономике и финансах»

Проверил: Доцент

\_\_\_\_\_\_\_\_\_\_\_\_\_\_/ Коровин А.М.

«\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_ 2023 г.

Автор работы:

студента группы КЭ – 314

\_\_\_\_\_\_\_\_\_\_\_\_/ Бухаров А.А.

«\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_2023 г.

# Цель метода аффинитивного анализа данных: В поиске скрытых связей и закономерностей

*«Сложные данные не всегда явно раскрывают свои секреты, но метод аффинитивного анализа данных помогает исследователям разгадывать их тайны».*

Современный мир буквально утопает в данных. От бизнеса и науки до медицины и социологии, огромные объемы информации создаются и собираются ежедневно. Однако за этой массой данных скрываются ценные знания и понимание. Цель метода аффинитивного анализа данных заключается в поиске и выявлении скрытых связей, закономерностей и структур в этом море информации.

**Суть Метода Аффинитивного Анализа Данных**

Метод аффинитивного анализа данных (Affinity Analysis) является мощным инструментом для обнаружения ассоциаций между данными. Основная его суть заключается в поиске паттернов и зависимостей в больших наборах данных. Важным применением этого метода является анализ товарных корзин в ритейле. Однако его применение не ограничивается этой областью и находит применение в различных сферах:

* **Маркетинг и продажи:** Помогает определить, какие товары чаще покупают вместе, что позволяет разрабатывать эффективные стратегии продаж и маркетинга.
* **Медицина:** Помогает выявлять связи между различными медицинскими симптомами и болезнями, что полезно для диагностики и лечения.
* **Финансы:** Позволяет выявлять зависимости между финансовыми инструментами и их влияние на рынок.
* **Научные исследования:** Помогает выявлять паттерны и закономерности в экспериментальных данных.

**Цель Метода Аффинитивного Анализа Данных**

Цель метода аффинитивного анализа данных заключается в выявлении ассоциаций и зависимостей между различными элементами данных. Он позволяет ответить на следующие вопросы:

1. **Какие элементы данных часто встречаются вместе?** Это помогает определить, какие товары можно рекомендовать вместе в ритейле или какие симптомы могут свидетельствовать о конкретных болезнях в медицинской области.
2. **Какие факторы оказывают влияние на определенные события или явления?** Например, в маркетинге, метод аффинитивного анализа данных может помочь определить, какие факторы влияют на успешность рекламных кампаний.
3. **Какие группы элементов данных образуют устойчивые паттерны?** Это может привести к выявлению новых знаний и пониманию сложных процессов.

**Заключение**

Метод аффинитивного анализа данных является важным инструментом для извлечения ценной информации из множества данных. Его цель заключается в обнаружении ассоциаций и зависимостей между данными, что позволяет делать информированные решения в различных областях. Этот метод продолжает развиваться и находить новые применения, делая его одним из ключевых инструментов анализа данных в современном мире.

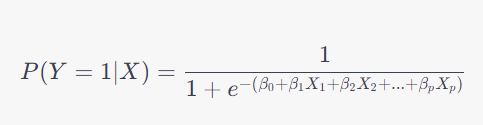
# Логистическая регрессия как специальная модификация регрессии

*«Логистическая регрессия - это мощный инструмент, который позволяет решать задачи классификации, прогнозирования вероятностей и анализа зависимости между переменными. Ее специфика и универсальность делают ее неотъемлемой частью анализа данных и машинного обучения».*

Логистическая регрессия - это один из наиболее распространенных методов машинного обучения, используемый в задачах классификации и прогнозирования вероятностей. Этот метод является специальной модификацией линейной регрессии, а его суть заключается в анализе связи между зависимой бинарной переменной и независимыми переменными.

**Сущность Логистической Регрессии**

Логистическая регрессия представляет собой математическую модель, которая использует логистическую функцию для прогнозирования вероятности принадлежности наблюдения к определенному классу. Основная идея заключается в том, что логистическая функция преобразует линейную комбинацию независимых переменных в вероятность. Модель логистической регрессии можно представить следующим образом:



Где:

* P(Y=1|X) - вероятность принадлежности к классу 1 при заданных значениях независимых переменных $X$.
* β0, β1, β2,, ... βp- коэффициенты регрессии, которые определяются в процессе обучения модели.
* X1, X2, ... , Xp - независимые переменные.

**Применение Логистической Регрессии**

Логистическая регрессия находит широкое применение в различных областях:

1. **Медицина:** Прогнозирование вероятности заболевания на основе клинических показателей.
2. **Финансы:** Оценка кредитоспособности клиентов и прогнозирование дефолтов.
3. **Маркетинг:** Анализ клиентской активности и прогнозирование вероятности оттока клиентов.
4. **Интернет-реклама:** Определение вероятности клика на рекламный баннер.
5. **Медиа:** Классификация текстов и анализ тональности.

**Преимущества и Недостатки Логистической Регрессии**

Преимущества:

1. **Простота и интерпретируемость:** Логистическая регрессия - это простая и понятная модель, которую легко интерпретировать.
2. **Эффективность:** В большинстве случаев логистическая регрессия показывает хорошую производительность, особенно при небольшом количестве признаков.
3. **Устойчивость к мультиколлинеарности:** Логистическая регрессия устойчива к наличию коррелированных признаков.

Недостатки:

1. **Линейная зависимость:** Логистическая регрессия не может улавливать сложные нелинейные зависимости между переменными.
2. **Плохо работает с несбалансированными данными:** В задачах с несбалансированными классами может потребоваться дополнительная настройка или использование других методов.

применение можно проиллюстрировать на примере задачи определения вероятности оттока клиентов в телекоммуникационной компании.

**Задача:** Представьте, что вы аналитик в телекоммуникационной компании, и вашей задачей является определение вероятности того, что клиент перейдет к конкуренту (отток клиентов). Для этого у вас есть исторические данные о клиентах, такие как продолжительность пользования услугами, сумма ежемесячных платежей, тип подписки и т. д. Вам нужно построить модель, которая предсказывает вероятность оттока для каждого клиента.

**Решение:** Для решения этой задачи можно использовать логистическую регрессию. Давайте посмотрим, как это работает:

1. **Подготовка данных:** Сначала необходимо подготовить данные. Это включает в себя очистку данных от нулевых значений, масштабирование признаков и разделение данных на обучающий и тестовый наборы.
2. **Построение модели:** Затем мы строим модель логистической регрессии, используя обучающий набор данных. Модель будет учиться на исторических данных и находить зависимость между признаками (например, продолжительностью пользования услугами) и целевой переменной (оттоком клиента).
3. **Оценка модели:** После построения модели мы оцениваем ее производительность с помощью тестового набора данных. Метрики, такие как точность, полнота и F-мера, могут быть использованы для оценки качества модели.
4. **Прогнозирование оттока:** Теперь, когда модель обучена и оценена, мы можем использовать ее для прогнозирования вероятности оттока для новых клиентов. Например, если новый клиент имеет высокую вероятность оттока, компания может предпринять меры для его удержания, например, предложив скидку на услуги или бонусы.

**Пример использования логистической регрессии:**

Предположим, у нас есть клиент, который пользуется услугами компании уже 3 года, его ежемесячный платеж составляет $70, у него подключена стандартная подписка. Мы хотим определить, склонен ли этот клиент к оттоку.

Модель логистической регрессии может предсказать вероятность оттока для этого клиента, например, она может предсказать, что вероятность оттока составляет 15%. Это предупреждение может помочь компании принять меры по удержанию клиента, например, предложив ему специальные условия или скидку на следующий месяц.

**Заключение**

Логистическая регрессия - это мощный инструмент в арсенале аналитика и исследователя данных. Она предоставляет возможность анализа и прогнозирования вероятностей, что делает ее неотъемлемой частью задач классификации и прогнозирования. Ее специфика и универсальность позволяют успешно применять этот метод в различных областях, что делает его одним из наиболее ценных инструментов в мире данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques. Morgan Kaufmann.
2. Tan, P.-N., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining. Pearson.
3. Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD international conference on Management of data (стр. 207-216).
4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer
5. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.
6. Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression. John Wiley & Sons.
7. Agresti, A. (2018). An Introduction to Categorical Data Analysis. Wiley.