

NTR Scoring – конструктор для создания скоринговых систем

Антон Мальков,
Исполнительный директор,
ООО «Лаборатория НТР» (NTR Lab)
malkov@ntrlab.ru
+7 095 9620386

NTR Scoring

- Система для кредитного скоринга (оценки кредитоспособности заемщика на основе демографических, исторических и ситуационных данных)
- Автор – компания «Лаборатория НТР», Москва
- Архитектура – клиент-серверный конструктор (решение)
- Подробная информация – сайт www.ntrlab.ru

История NTR Scoring

- 2000-2001 – заказная разработка системы кредитного скоринга для одного из лидеров рынка потребительского кредитования
- 2004 – продукт NTR Scoring v.2

NTR Lab: подход к кредитному скорингу



- Кредитный скоринг: широко применяется с 1966 года
- Классические методы опираются на кредитную историю
- Российская (и вообще по СНГ) ситуация – отсутствие не только кредитной истории но, зачастую, и адекватно верифицируемых доходов

Проблемы потребительского кредитования



- Непрозрачность расходов клиентов
- Непрозрачность доходов клиентов
- Отсутствие истории взаимоотношений с клиентом
- Отсутствие гарантий доходов клиента на срок кредитования
- Необходимость выдавать кредиты на точке продажи очень быстро

NTR Scoring – используемая информация



- Демографическая информация - это анкетная информация о клиенте.
- Ситуационная информация - информация о том за каким кредитом, в какое место и время пришел клиент. В случае револьверного кредитования такая информация отсутствует.
- Историческая информация - информация об истории финансовых операций с клиентом. Пока что в большинстве случаев такая информация отсутствует.

NTR Scoring – общий подход к выбору и построению методов



- Адаптивные методы, основанные на расширенной демографии и анализе точности данных. Здесь – две идеи:
 - Взять сначала как можно больше анкетных и ситуационных данных о клиенте. В дальнейшем те пункты анкеты, которые не влияют на кредитный риск, отбросить
 - Анализировать достоверность данных, представляемых клиентом. Если есть проблемы, отказывать в кредите

NTR Scoring – общая архитектура



- Возможность конструировать скоринговую систему на основе голосования нескольких скоринговых моделей
- Возможность использования системы с собственным front-endом или интеграция с системой банка - заказчика на основе веб-сервисов

Функциональность системы

Модули

- Проверка заявки
- Кредитный скоринг
- Фронт-офис

Проверка заявки

- на полноту и достаточность информации (заявка проверяется сама по себе)
- на наличие информации о клиенте в черном списке
- выполнение проверки информации на внешние условия (по демографическим базам данных)
- Проверка массива заявок (проверки типа не взяла ли кредит жена или на совпадение телефонов)
- Примерно по 40 проверок каждого типа из трех (черный список – очевидно, одна проверка)

Логический контроль

- Проверка каждого поля на допустимость соответствующего значения
- Проверка совокупности полей на непротиворечивость

Результатом логического контроля является список полей заявки требующих уточнения.

Поиск совпадений

- Каждая вновь вводимая в клиентскую БД заявка на получение кредита сравнивается со всеми ранее введенными заявками на предмет поиска совпадений в полях, аналогичных по содержанию. В случае успеха поиска заявка снабжается списком всех найденных совпадений и указателей на ранее введенные заявки, где обнаружилось совпадения. В процессе принятия кредитного решения по вновь поступившей заявке операторы имеют возможность:
- увидеть полный список совпадений, ассоциированный с данной заявкой;
- просмотреть любую из заявок, с данными которых обнаружилось совпадения.

Кредитный скоринг

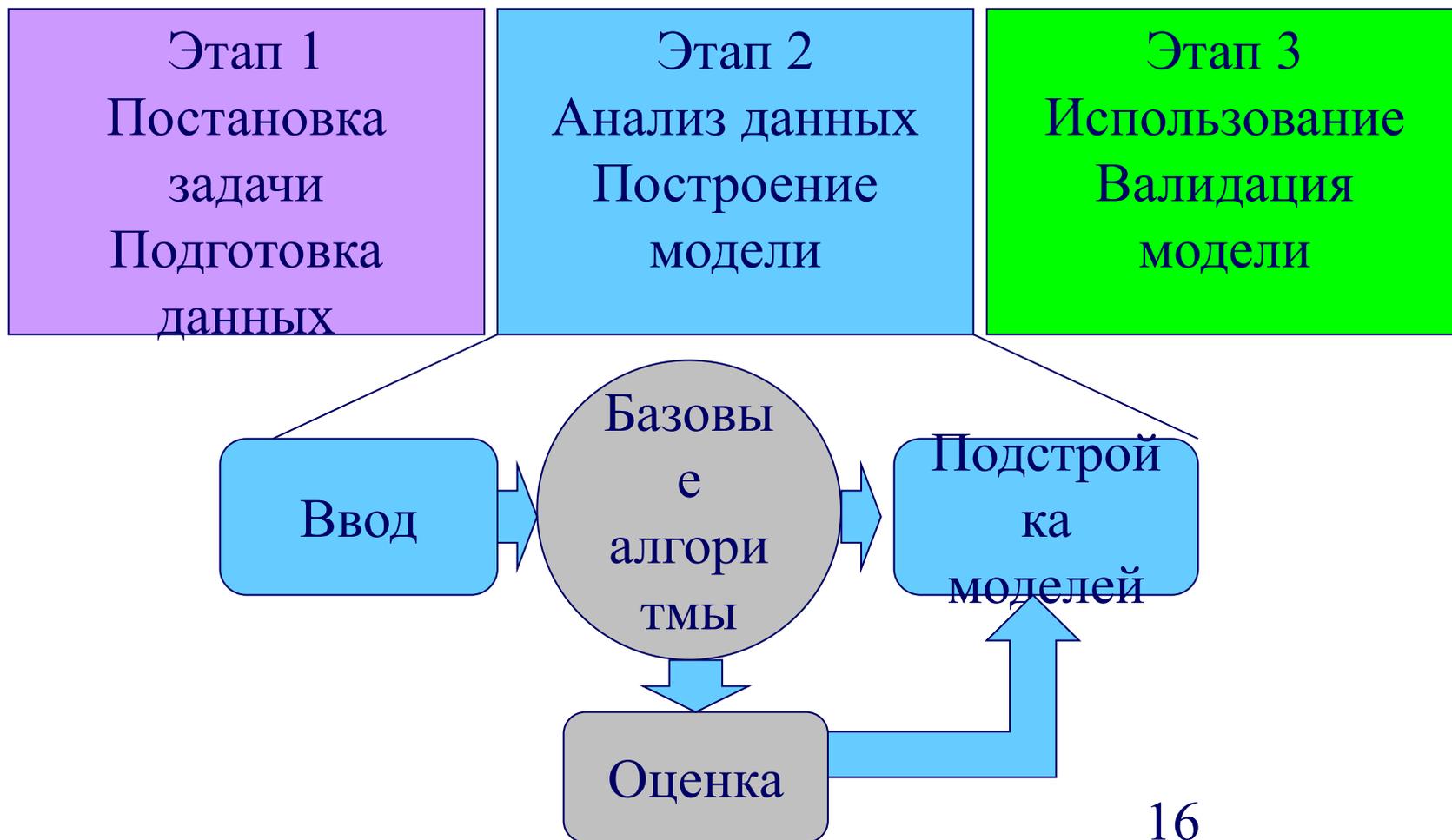
- Ключевое предположение
 - Прошлый опыт может быть использован для предсказания платежеспособности заемщика
- Два основных типа моделей
 - Основанные на правилах (отображение существующей кредитной политики банка)
 - Статистические (основанные на анализе данных о клиентах банка)

Модели, основанные на правилах, в NTR Scoring

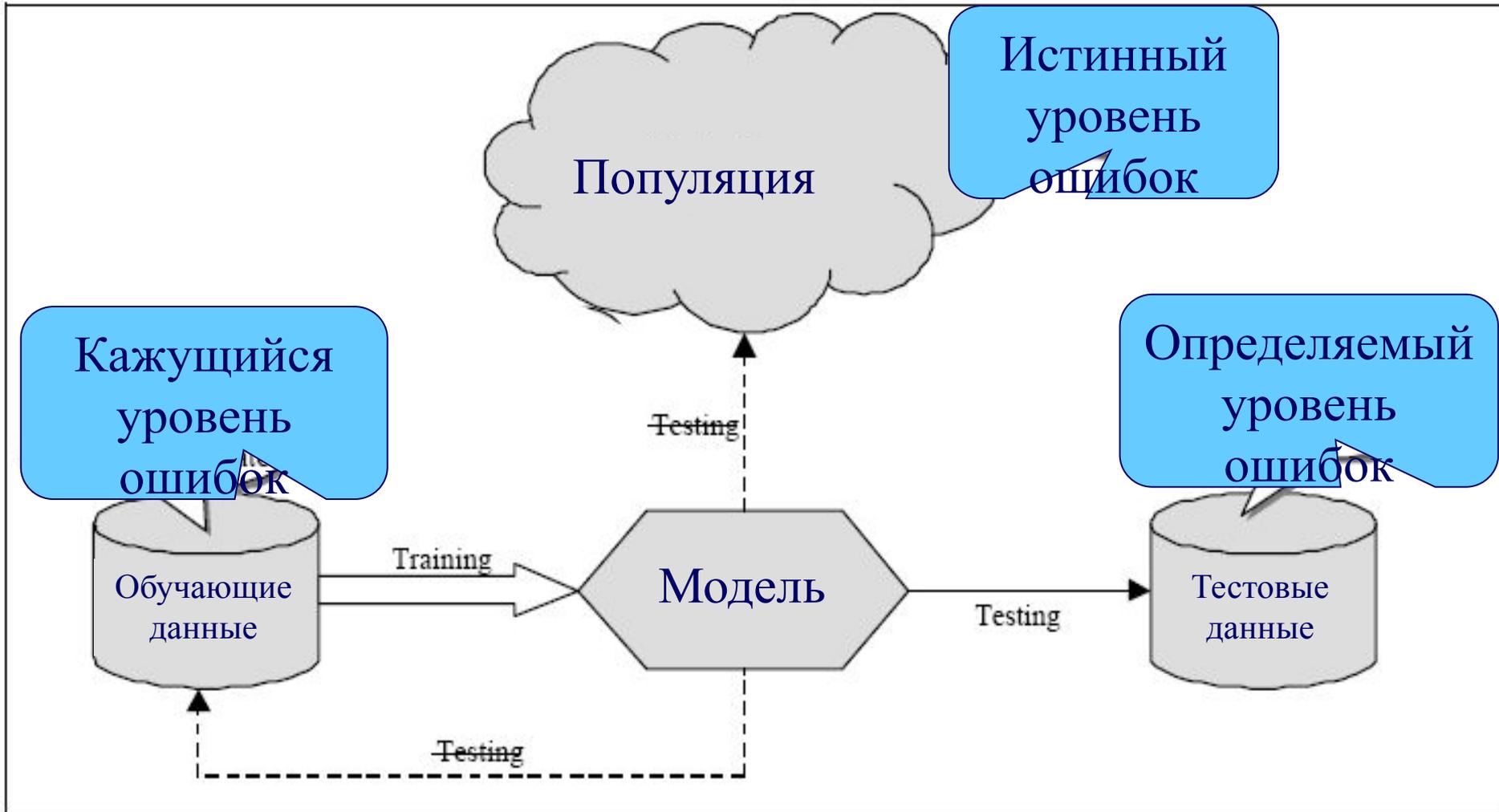


- **Линейная скоринговая модель**
 - Для каждого признака определяется его вес, и клиенту выставляется оценка по каждому признаку. Оценки суммируются с учетом весов.
- **Дерево решений**

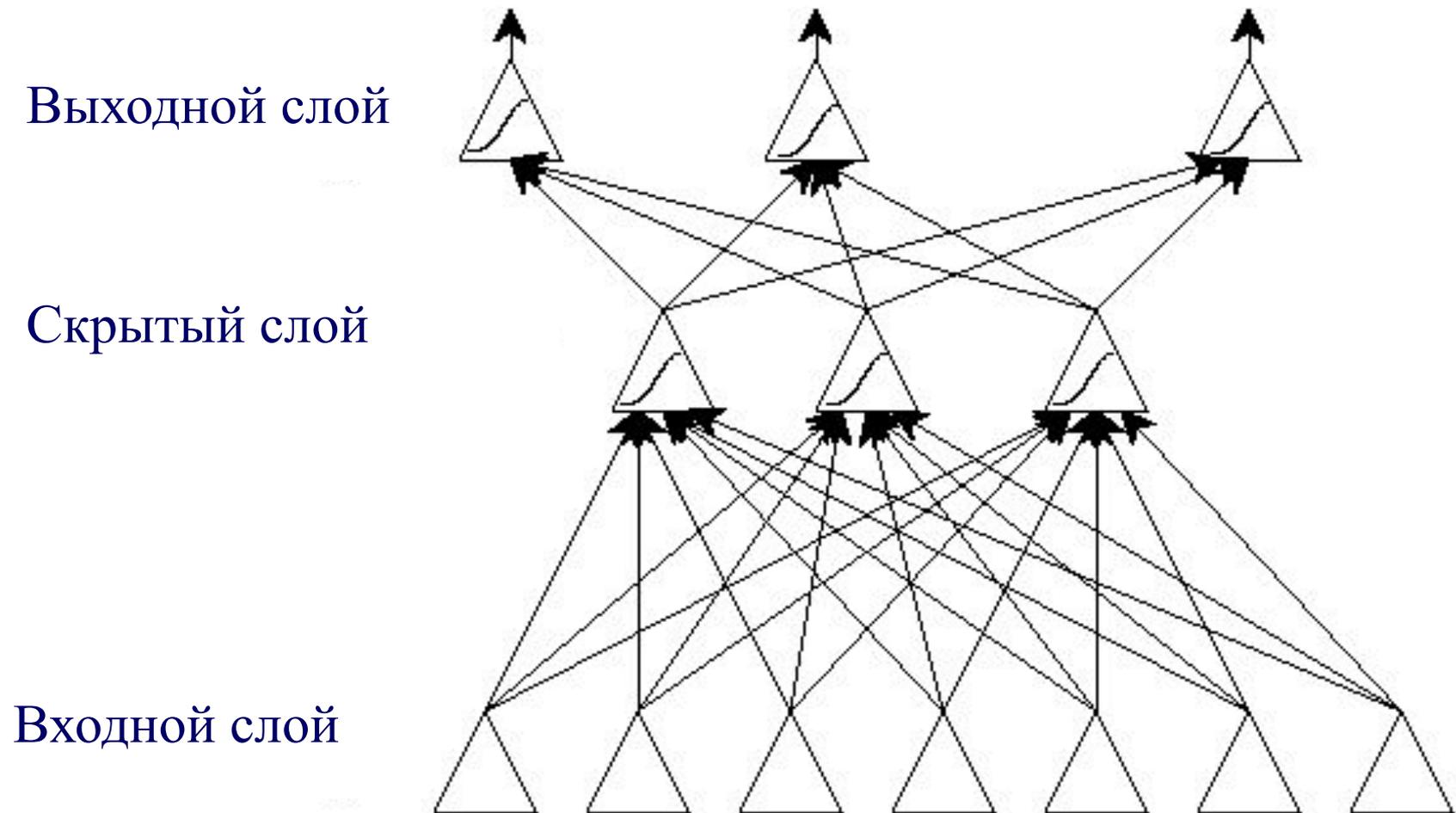
Общая схема построения статистических моделей



Обучение моделей

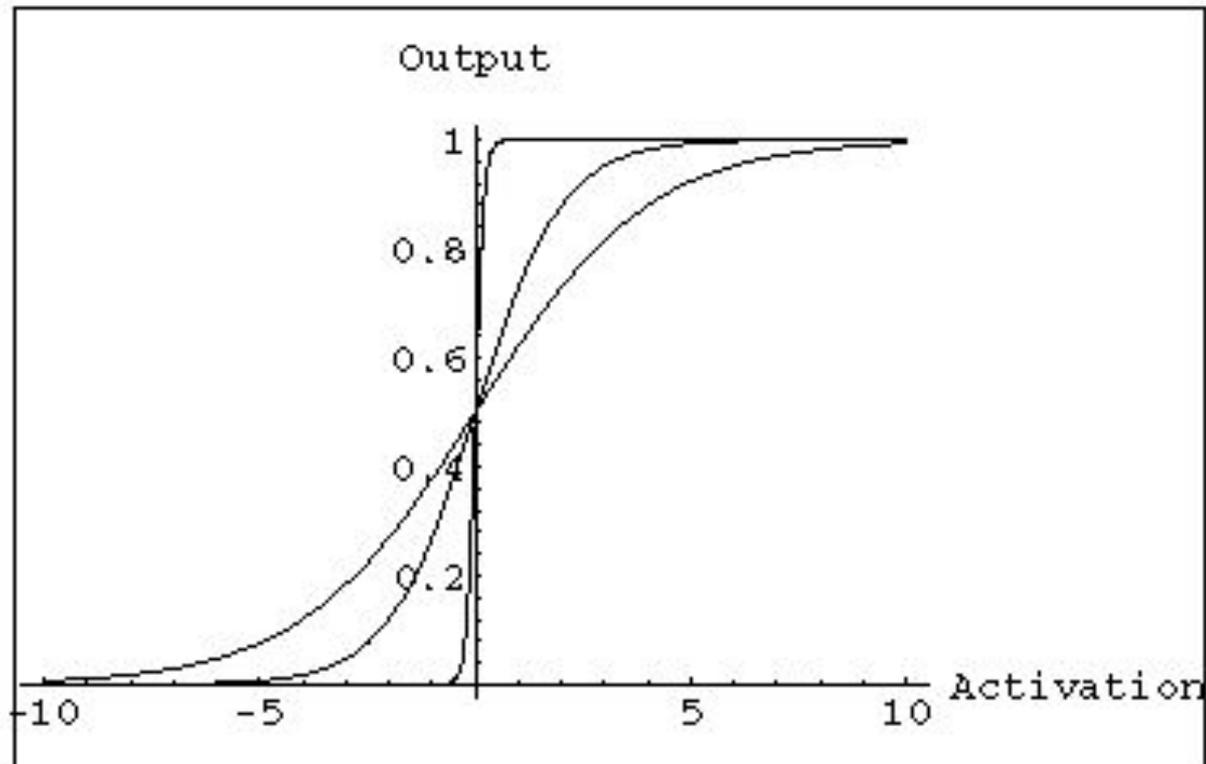


Нейронные сети



Передаточная функция

- $Y = 1 / (1 + \exp(-k(\sum W_{in} * X_{in})))$



Обучение

- Прямой ход:
 - Вычисляются результаты для примера и ошибки.
- Обратный ход (обратное распространение ошибки):
 - Ошибки на выходе используются для изменения весов на выходных нейронах. Потом вычисляются ошибки на нейронах скрытого слоя, и на их основе изменяются веса.

k ближайших соседей (k-NN)

- Общий принцип – обучение по примерам. Запоминаются обучающие примеры и для принятия решения по новому примеру ищутся наиболее похожие обучающие примеры
- k-NN – выбираются k ближайших соседей (расстояние может измеряться по-разному). Пример относится к тому классу, соседей которого класса больше (возможно, с весом, обратно зависящим от расстояния)

Principal Component Analysis

- Тоже метод обучения по примерам
- Данные переводятся в двумерную плоскость, соответствующую главным компонентам (старшим сингулярным векторам) матрицы распределения обучающих примеров
- Далее – k-NN
- Позволяет отфильтровать шум в выборке

Локально взвешенная регрессия (обучение по примерам)



- Тоже метод обучения по примерам
- В окрестности примера строится линейная регрессионная модель, дающая непрерывную вероятность принадлежности к классам
- Образцы для модели взвешиваются с использованием обратно зависящего от расстояния веса

Линейный дискриминантный анализ (1)



- Предположим, что нам надо классифицировать клиентов по двум классам c_1 и c_2
- X – вектор анкеты клиента
- L_{ij} – ожидаемая цена ошибки классификации
- $\Pr(X)$ – Вероятность реализации X .
- $\Pr(X|c_i)$ – условная вероятность X в классе c_i .

Линейный дискриминантный анализ (2)



- Применим для классификации байесовское правило: отнести X к классу c_1 , если

$$\ln \frac{\Pr(X | c_1)}{\Pr(X | c_2)} > \ln \frac{L_{21} \Pr(c_2)}{L_{12} \Pr(c_1)}$$

- Предположим, что распределение векторов в каждом классе $f(\mathbf{X}|c_i)$ соответствует нормальному гауссовскому распределению. Тогда

$$\hat{\Pr}(X | c_i) = f(\mathbf{X} | c_i) |_{\mathbf{X}=X} = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (X - \mu_i)^T \Sigma^{-1} (X - \mu_i)\right], \quad i=1,2$$

где μ_i — средний вектор для класса i , а Σ — матрица ковариации

Линейный дискриминантный анализ (3)



- Таким образом, правило классификации:

$$X^T \Sigma^{-1} (\mu_1 - \mu_2) > \frac{1}{2} (\mu_1 + \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2) + \ln \frac{L_{21} * \Pr(c_2)}{L_{12} * \Pr(c_1)}$$

- μ_i и Σ могут быть получены из исторических данных
- основные проблемы – оценка $(\Pr(c_2)/\Pr(c_1))$ (точность зависит от представительности выборки) и (L_{21}/L_{12})
- Можно заменить (L_{21}/L_{12}) и $(\Pr(c_2)/\Pr(c_1))$ константой отсечки и подбирать ее

Функции фронт-офиса

- Ведение жизненного цикла кредита
- Управление информацией о клиентах
- Управление информацией о Продуктах
- Управление скорингом

Ведение жизненного цикла кредита



- Открытие кредита
- Выпуск пластиковой карты
- Отслеживание погашения кредита с учетом грейс-периода
- Возобновление кредита
- Управление досрочным закрытием

Управление информацией о клиентах



- создание единой базы данных по клиентам Банка, зарегистрированных в рамках Системы;
- автоматизация процессов регистрации и обработки заявок клиентов Банка на предоставление Продуктов в рамках Системы;
- автоматизация процесса принятия решения о кредитоспособности клиентов на основе процедуры скоринга;
- обеспечение целостности информации по клиентам в Системе;
- накопление кредитной истории клиентов Банка.

Управление информацией о Продуктах



- автоматизация процедур управления продуктами;
- обеспечение целостности информации по кредитам в Системе;
- получение статистической и аналитической информации по использованию продуктов Банка.

Управление скорингом

- анализ истории предоставления кредитов
- расчет и перерасчет скоринговых коэффициентов

Функции управления клиентами



- регистрация и ведение заявок клиентов на предоставление Продукта;
- выполнение проверок зарегистрированных заявок;
- выполнение расчета кредитного рейтинга клиента (скоринг);
- регистрация и ведение информации о клиентах;
- управление статусами клиентов;
- сбор информации о клиентах от других модулей Системы;
- предоставление информации о клиентах другим модулям Системы;
- регистрация событий, связанных с жизненным циклом клиента.

Функции управления продуктами



- регистрация и ведение информации о кредитах;
- регистрация событий, связанных с жизненным циклом кредита;
- управление статусами кредитов;
- сбор информации о кредитах от других модулей Системы;
- предоставление информации о кредитах другим модулям Системы.

Вопросы, предложения?



Антон Мальков,
Исполнительный директор,
ООО «Лаборатория НТР» (NTR Lab)
malkov@ntrlab.ru
+7 095 9620386