

Метрики качества прогноза

Д. Звежинский, 19.12.2018

Dmitry.Zvezhinsky@sas.com

Ошибка прогноза

- Измерение и изучение ошибки прогноза – базовый инструмент в статистике:
 - Построение модели: оптимизация ошибки.
 - Анализ построенной модели: в каких случаях вклад в ошибку прогнозирования больше?
 - Анализ сложности построенной модели: есть ли переобучение (на отложенной выборке)?
 - Анализ построенной модели: в регрессионных методах – проверка предположений для анализа.
 - Выбор наилучшей модели из нескольких.

Ошибка прогноза

Что меняется, если переходим к временным рядам?

- Прогноз выдаётся на различные горизонты (периоды будущего). Это - параметр прогноза.
- При кросс-валидации обучающая и проверочная выборки делятся по времени (известное прошлое и неизвестное будущее)
- Если ряд один – для подбора моделей и оценки точности можно пользоваться разными критериями согласия (Statistics of fit).
- Если рядов много – интереснее смотреть на агрегат по всем временным рядам, некоторую «ошибку всего прогноза» :

Ошибка всего прогноза = $\sum f(\text{прогноз}_{i,j} - \text{факт}_{i,j})$

i=периоды ряда, j=ряды

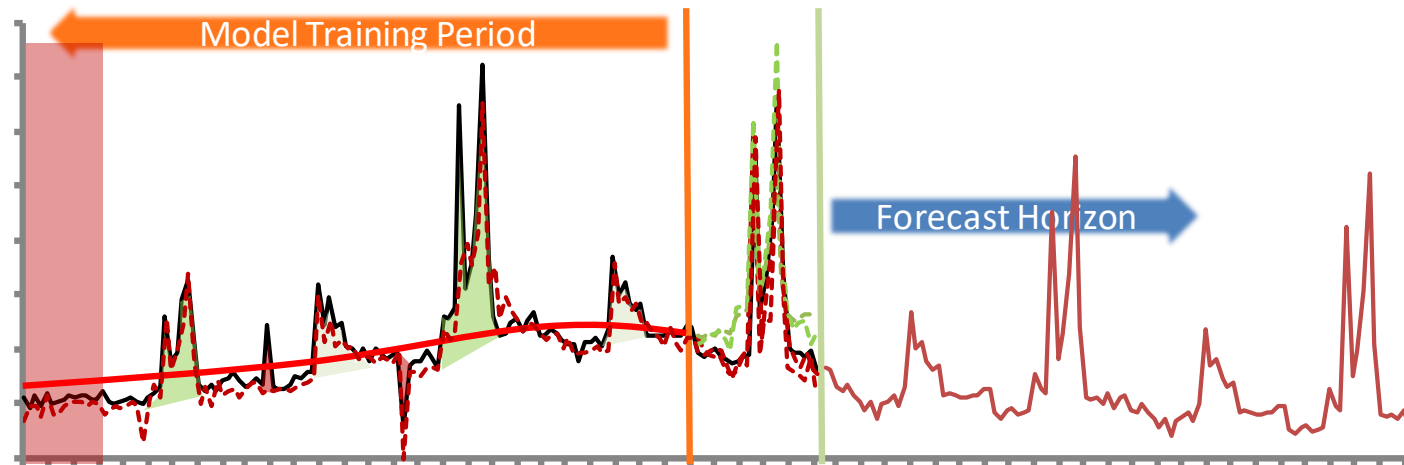
Критерии согласия (statistics of fit)

- В SAS для подбора моделей временных рядов есть 49 критериев согласия:
- <https://support.sas.com/documentation/onlinedoc/forecast/14.2/fsug.pdf> (p.385)
- При разных критериях согласия будут лучшими разные модели.
- Обратите внимание, что эти критерии не используются для подбора коэффициентов одной модели, они используются для выбора между несколькими моделями для одного временного ряда.
- См следующий слайд.

Прогнозирование

90% временных рядов прогнозируются по следующей схеме

6 Прогнозирование временных рядов с помощью факторных моделей



Ошибка

13.64%

11.05%

8.58%

Модель 1: ARIMA, $f(\text{история, выбросы, цена, промоакции, остатки, НГ и другие праздники, взаимовлияние})$

Модель 2: Сезонное экспоненциальное сглаживание, $f(\text{история, сезонность})$

Модель 3: ARIMA, $f(\text{история, сезонность, выбросы, цена, промоакции, НГ и другие праздники})$

■ Выбросы



Методика расчета общей ошибки

$$\text{Ошибка всего прогноза} = \sum_{i,j} f(\text{прогноз}_{i,j} - \text{факт}_{i,j})$$

Выбор функции f – это только первый шаг к расчету общей ошибки. Что нужно решить дальше?

1. Какие периоды используются в расчете ошибки?

- прогноз и факт по дням? по неделям?
- в расчет могут входить только периоды с определёнными характеристиками, например, с регулярными продажами, без дефицита и выбросов.
- какие горизонты включаются в расчет ошибки?

Метрики ошибки

$$\text{Ошибка всего прогноза} = \sum_{i,j} f(\text{прогноз}_{i,j} - \text{факт}_{i,j})$$

2. Какие ряды используются в расчете?

- учитываются ли ряды без достаточной истории?
- ряды с нестабильным спросом?
- ряды, которые дают относительно низкие обороты? (группа С)

3. Детализация ошибки: отдельно для периодов с промо и без промо, отдельно для разных групп товаров/магазинов

Метрика ошибки: влияние на бизнес-результат

$$MAE = \sum |F - A|$$

- + Вес Перепрогноза = Вес Недопрогноза
- - Не учитывает вес товара в обороте компании

$$WAPE = \frac{\sum |F - A|}{\sum |A|}$$

- + Учитывает вес товара в обороте компании
- + Привычен ТС5 (ранее использовался в отчётах)
- + Выявляет аномально большие прогнозы
- - Приводит к недопрогнозу

$$SWAPE = \frac{\sum \left(\frac{|F - A|}{F + A} * A \right)}{\sum |A|}$$

- Не занижает прогноз спроса
- Вес Перепрогноза = Вес Недопрогноза
- Учитывает вес товара в обороте компании
- Значение от 0 до 1 => Простая связь с точностью: Accuracy = 1-SWAPЕ
- Похож на WAPE (так же интерпретируется, значения того же масштаба)
- - Не чувствителен к аномальным (в 100 раз) отклонениям прогноза от цели

Метрика ошибки: разница в ответах для WARE, SWARE, MSE

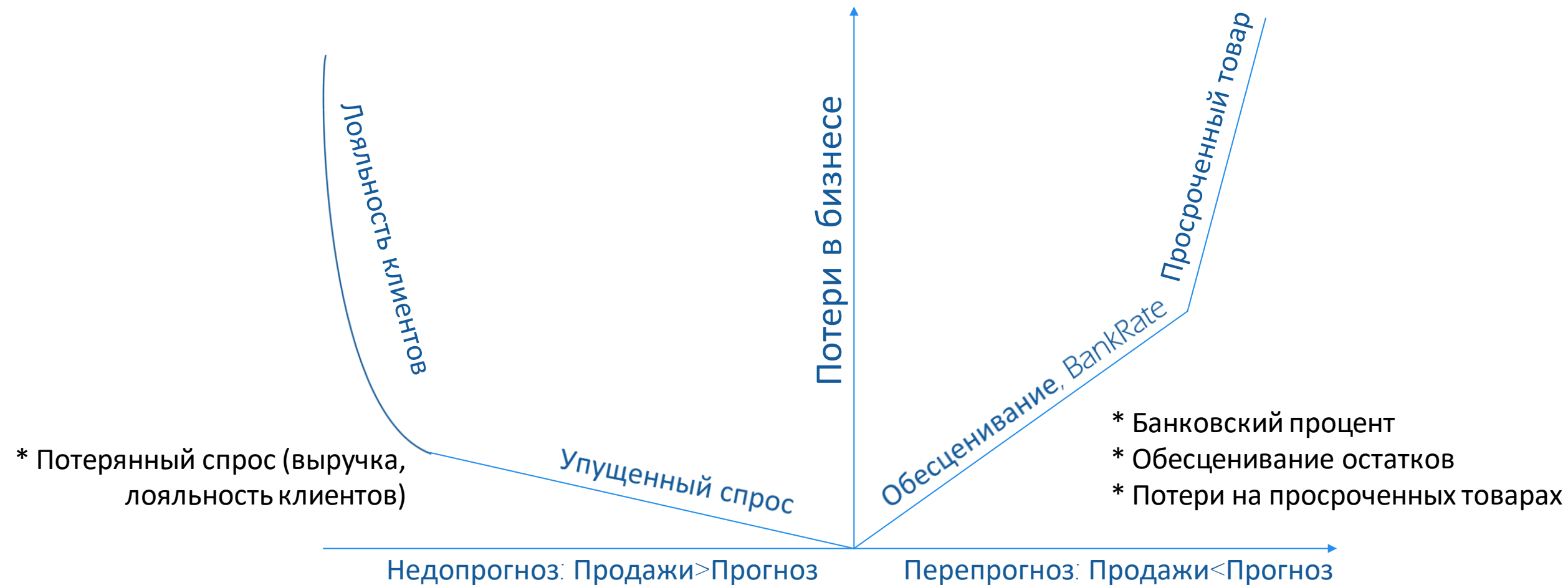


Метрики ошибки: пост-обработка прогноза для минимизации ошибки

Поведение Ошибки WARE

SAS	Занулили дневной прогноз* при недельном <=				
	3	4	5	6	7
69,1	67,4	66,7	66,4	66,3	66,4
72,1	69,4	68,7	68,3	68,2	68,3
74,6	70,7	70,0	69,6	69,5	69,5
75,3	71,3	70,8	70,5	70,5	70,7

Функция потерь при прогнозировании



KPI в Retail

На практике компании отслеживают не «точность всего прогноза», а другие метрики (KPI), измеряемые в штуках или деньгах. Эти метрики неочевидно связаны с ошибкой прогноза, скорее со всей цепочкой поставок (исполнение заказов, оптимизация остатков, логистика ...)



Показатели точности прогноза

$$MSE = (\text{Прогноз} - \text{Продажи})^2$$

$$MAE = |\text{Прогноз} - \text{Продажи}|$$

$$MAPE = \frac{|\text{Прогноз} - \text{Продажи}|}{\text{Продажи}}$$

Примеры KPI

- Оборачиваемость = Продажи / Запасы (чем больше – тем лучше)
- Можно брать разные запасы (в магазинах, на складах, товар в пути).
- Запасы за неделю / Продажи за неделю = сколько недель товар лежит в сети перед продажей.
- Уменьшение срока продажи на 1 день -> от 5 до нескольких десятков млн \$ в год.

Домашнее задание

На данных «утконоса»:

- Реализовать наивные прогнозы *спроса* (будущее значение = предыдущему или будущее значение = скользящему среднему за прошлые периоды).
- Измерить точность наивного прогноза *спроса* на горизонты 1..3
- В качестве метрики возьмём MAE
- Нужные разрезы: горизонт – категория товаров – все магазины.
- Из ошибки НЕ выбрасываем дни дефицита, промо.