



Data Mining in Action

Лекция 5. Метрики качества



Образовательный партнер курса



misis.ru

Генеральный партнер курса



jet.su

Партнеры курса



raiffeisen-digital.ru



academy.yandex.ru

На прошлой лекции

- Решающие деревья
- Ансамбли деревьев
- Общие идеи построения ансамблей
- Извлечение и простые преобразования признаков
- Отбор признаков

Немного мотивации: топ ошибок в индустрии

1. Постановка задачи отсутствует или неправильная (например, метрику вообще выбрали случайно)
2. A/B тест не проводится или не валиден
3. Утечка и переобучение

Субъективный топ причин

1. Безответственность: «и так сойдет»
2. Невнимательность, особенно в период «авралов»
3. Нехватка экспертизы: незнание, что вопросы, которые мы обсудим на этой лекции, существуют и важны

План

1. Метрики в задачах регрессии

2. Метрики в задачах классификации

3. Пример выбора метрики

1. Метрики в задачах регрессии

Функционал ошибки (loss)

- MAE
- RMSE
- MAPE
- SMAPE
- logloss

MEAN AVERAGE ERROR

- Отклонение прогноза от исходного значения
- Усредненное по всем наблюдениям

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

ROOT MEAN SQUARED ERROR

- Корень из среднего квадратичного отклонения прогноза от исходного значения
- Сильнее штрафует за бОльшие по модулю отклонения

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

MEAN AVERAGE PERCENTAGE ERROR

- Ошибка прогнозирования оценивается в процентах

$$M = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

SYMMETRIC MEAN AVERAGE PERCENTAGE ERROR

- Ошибка оценивается в процентах

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

SYMMETRIC MEAN AVERAGE PERCENTAGE ERROR

- Ошибка оценивается в процентах

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{|A_t| + |F_t|}$$

SYMMETRIC MEAN AVERAGE PERCENTAGE ERROR

- По-разному штрафует за перепрогнозирование и недопрогнозирование

- Перепрогнозирование:

$$A_t = 100, F_t = 110 \sim \text{SMAPE} = 4.76\%$$

- Недопрогнозирование:

$$A_t = 100, F_t = 90 \sim \text{SMAPE} = 5.26\%$$

Log Loss

- Логарифмическая ошибка
- Хорошо оценивает вероятность

$$\text{LogLoss} = -\frac{1}{n} \sum_{i=1}^n \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]$$

Почему Log Loss так выглядит

Пусть $p_i = P(y_i = 1|x_i)$, тогда $1 - p_i = P(y_i = 0|x_i)$

Почему Log Loss так выглядит

Пусть $p_i = P(y_i = 1|x_i)$, тогда $1 - p_i = P(y_i = 0|x_i)$

Теперь заметим, что выражение $p_i^{y_i}(1 - p_i)^{(1-y_i)}$ - просто запись вероятности того класса, к которому x_i фактически принадлежит

Почему Log Loss так выглядит

Пусть $p_i = P(y_i = 1|x_i)$, тогда $1 - p_i = P(y_i = 0|x_i)$

Теперь заметим, что выражение $p_i^{y_i}(1 - p_i)^{(1-y_i)}$ - просто запись вероятности того класса, к которому x_i фактически принадлежит

Произведение вероятностей фактических классов объектов из выборки - правдоподобие выборки:

$$\prod_{i=1}^n p_i^{y_i}(1 - p_i)^{(1-y_i)}$$

Почему Log Loss так выглядит

Пусть $p_i = P(y_i = 1|x_i)$, тогда $1 - p_i = P(y_i = 0|x_i)$

Теперь заметим, что выражение $p_i^{y_i}(1 - p_i)^{(1-y_i)}$ - просто запись вероятности того класса, к которому x_i фактически принадлежит

Произведение вероятностей фактических классов объектов из выборки - правдоподобие выборки:

$$\prod_{i=1}^n p_i^{y_i}(1 - p_i)^{(1-y_i)}$$

Если взять логарифм и умножить на -1 - получим log loss

Log Loss константного прогноза

Рассмотрим выборку из n объектов с одинаковыми векторами признаков x , на pn из которых таргет равен 1, а на остальных – 0.

Log Loss константного прогноза

Рассмотрим выборку из n объектов с одинаковыми векторами признаков x , на pn из которых таргет равен 1, а на остальных – 0.

Пусть $a(x) = c$, тогда log loss минимален при:

$$\left(\sum_{i=1}^n y_i \ln c + (1 - y_i) \ln(1 - c) \right)'_c = 0$$

Log Loss константного прогноза

$$\left(\sum_{i=1}^n y_i \ln c + (1 - y_i) \ln(1 - c) \right)'_c = 0$$

$$\frac{pn}{c} - \frac{n - pn}{1 - c} = 0$$

$$pn - cpn = cn - cpn$$

$$pn = cn$$

$$c = p$$

История про MAE вместо log loss

- Заказчик очень хотел, чтобы алгоритм оценивал вероятности в задаче бинарной классификации
- Немного знал про функции потерь
- Просил решать задачу регрессии на ответах 0 и 1 оптимизируя MAE, думал ответы будут между 0 и 1
- Ответы получились только 0 и 1

Упражнение

1. Показать, что если вместо $\log \text{loss}$ оптимизировать MAE в задаче с ответами 0 и 1, прогноз алгоритма будет округляться к 0 или к 1
2. Показать, что константный прогноз в регрессии, оптимизирующий MSE – среднее значение таргетов

2. Метрики в задачах классификации

Метрики качества

- Accuracy
- Precision
- Recall
- F-measure
- ROC-AUC

Accuracy

Доля правильных ответов при классификации

Accuracy

Доля правильных ответов при классификации

target: 1 0 1 0 0 0 0 1 0 0

Accuracy

Доля правильных ответов при классификации

target: 1 0 1 0 0 0 0 1 0 0

predicted: 0 0 1 0 0 0 0 1 1 0

Accuracy

Доля правильных ответов при классификации

target: 1 0 1 0 0 0 0 1 0 0

predicted: 0 0 1 0 0 0 0 1 1 0

Accuracy

Доля правильных ответов при классификации

target: 1 0 1 0 0 0 0 1 0 0

predicted: 0 0 1 0 0 0 0 1 1 0

accuracy = 8/10 = 0.8

Метрики качества

- Accuracy
- Precision
- Recall
- F-measure
- ROC-AUC

Precision & Recall

- Precision – точность
- Recall - полнота

Сбитые самолеты



Сбитые самолеты



$y = (0\ 0\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1)$

$\hat{y} = (0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1)$

Precision

Precision – точность выстрелов:

Количество сбитых самолётов

Количество выстрелов

$$y = (0\ 0\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1)$$

$$\hat{y} = (0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1)$$



Recall

Recall – «полнота» сбивания самолетов:

Количество сбитых самолётов

Общее количество самолётов

$$y = (0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1)$$

$$\hat{y} = (0 \ 1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1 \ 0 \ 1)$$



Обычно объясняется так:

		Actual Class	
		Yes	No
Predicted Class	Yes	T True P ositive	F alse P ositive
	No	F alse N egative	T True N egative

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

F-measure (F-score, F1)

- Среднее гармоническое между precision и recall:

$$F1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

- Значение F-measure ближе к меньшему из precision и recall

Метрики качества

- Accuracy
- Precision
- Recall
- F-measure
- ROC-AUC

ROC-AUC

- Применяется для оценки «вероятностной» классификации*
- «Качество» ранжирования объектов по вероятности принадлежности к целевому классу
- Доля «правильно» отранжированных пар
- Вероятность встретить объект целевого класса раньше, чем объект нецелевого класса

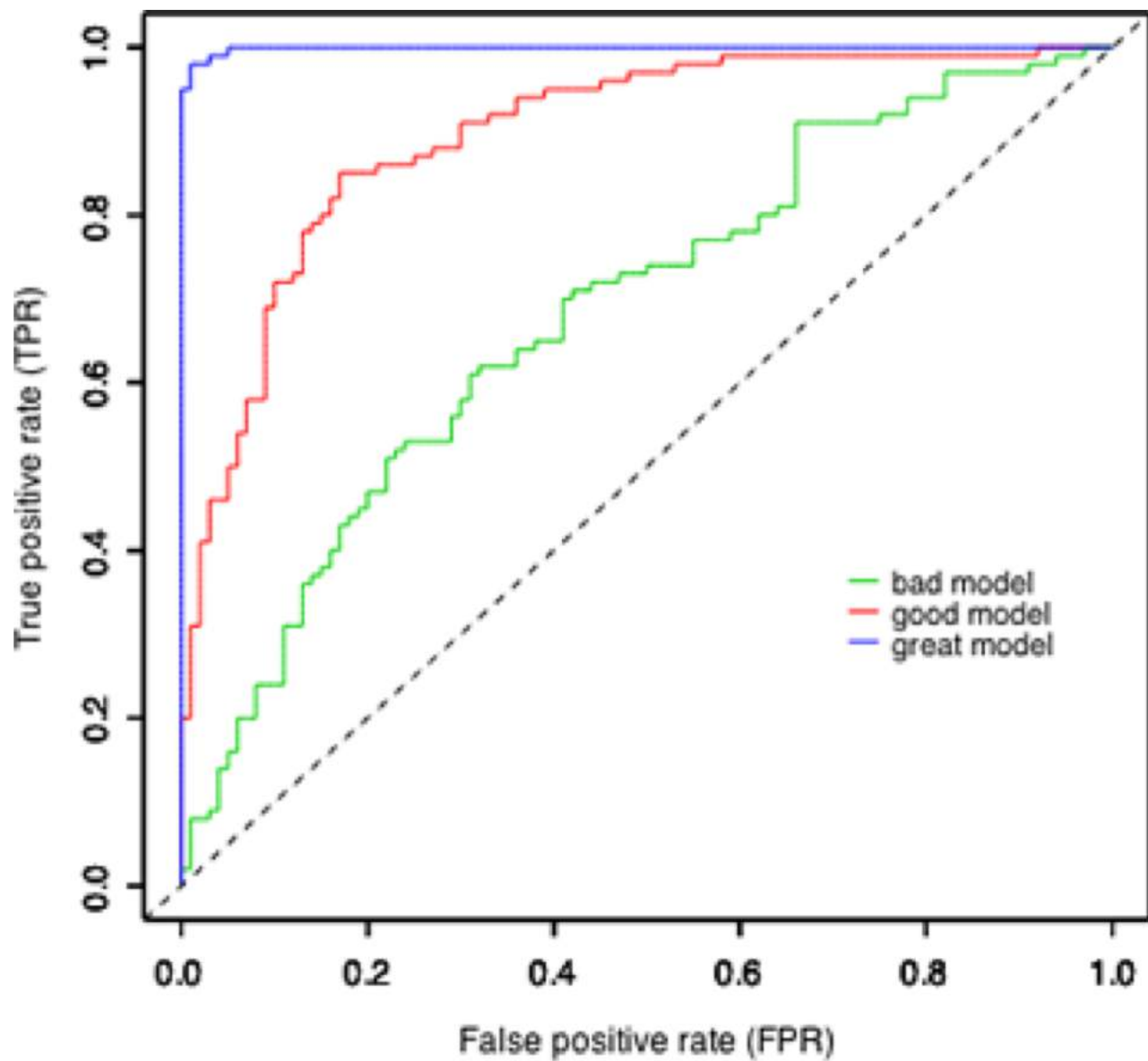
ROC

		Actual Class	
		Yes	No
Predicted Class	Yes	True Positive	False Positive
	No	False Negative	True Negative

$$TPR = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

$$FPR = \frac{\text{False positives}}{\text{False positives} + \text{True negatives}}.$$

ROC



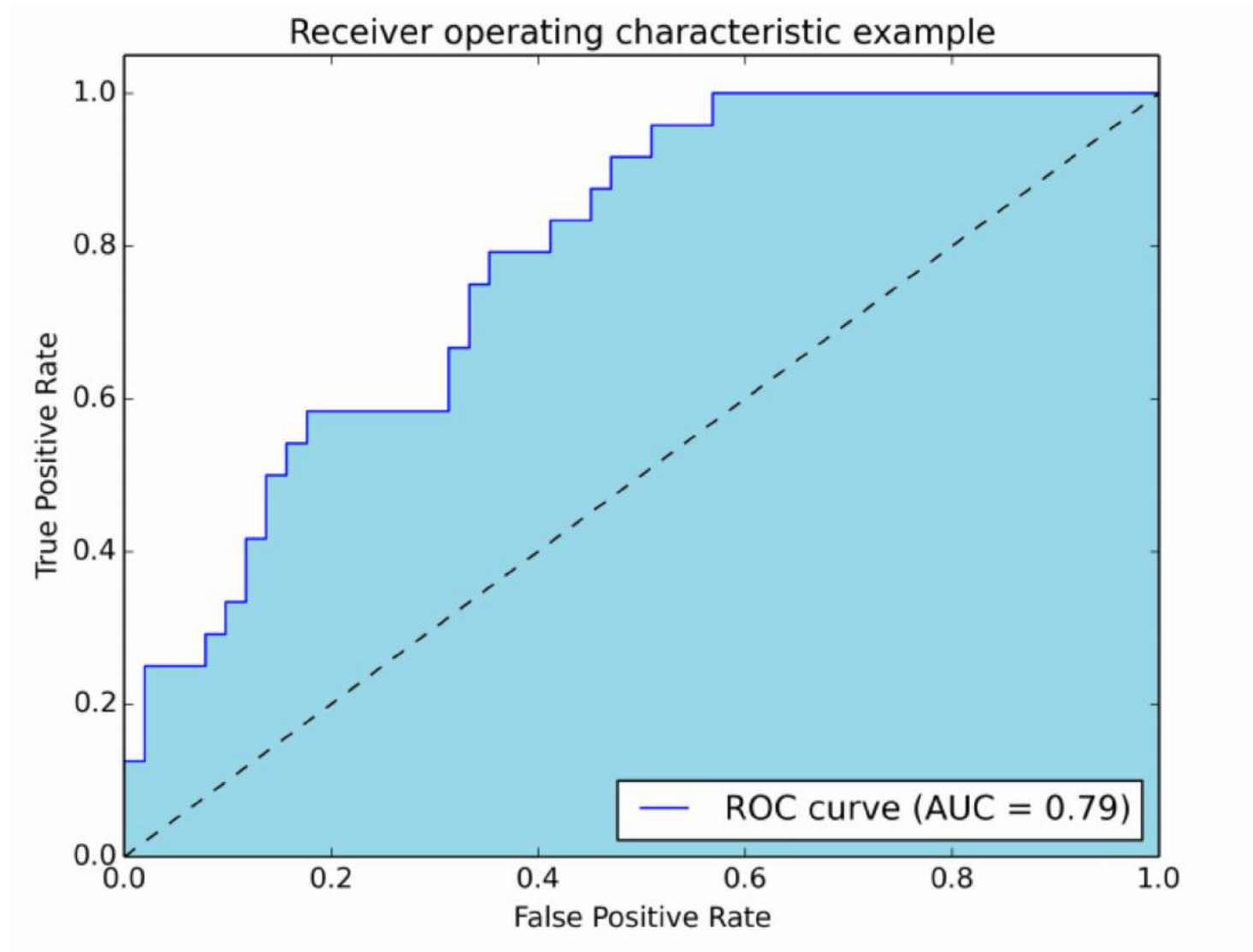
ROC

- Как оценить кривую численно?

ROC-AUC

- Как оценить кривую численно?
- Измерить площадь под кривой – area under the curve!

ROC-AUC

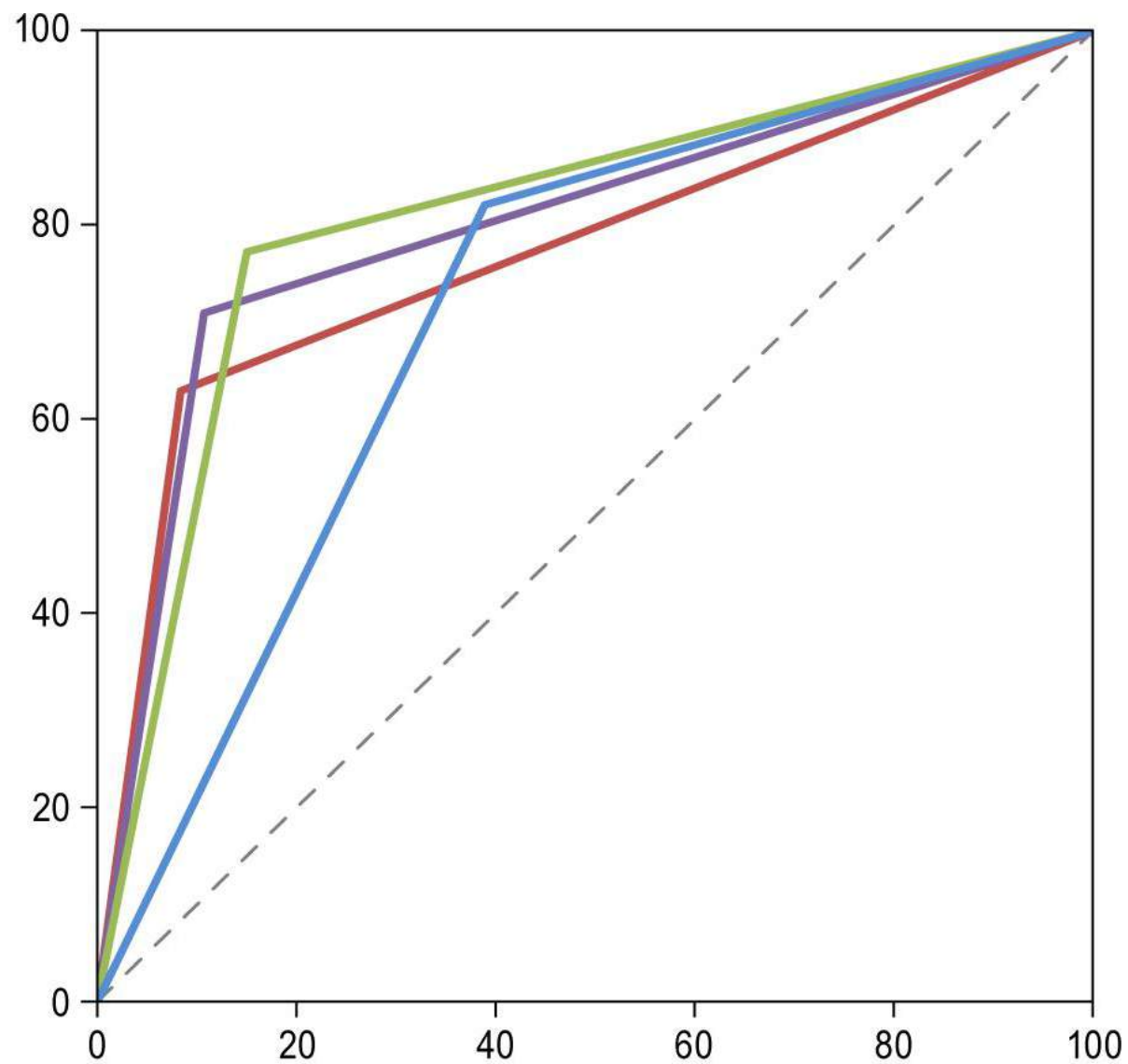


ROC-AUC по-простому

Рассмотрим всевозможные пары объектов из выборки. ROC-AUC – доля тех пар, которые алгоритм отранжировал правильно.

История про ROC-AUC по 0 и 1

История про ROC-AUC по 0 и 1



Упражнение

- Показать, что треугольный ROC-AUC для константного ответа равен 0.5
- Показать, что треугольный ROC-AUC для случайного ответа 0 или 1 (с любой вероятностью ответа 1) – тоже равен 0.5
- Показать, что обычный ROC-AUC для случайных ответов из равномерного распределения на $[0, 1]$ равен 0.5

Дополнительные материалы

Рассказ про ROC-AUC в блоге Александра Дьяконова:

<https://dyakonov.org/2017/07/28/auc-roc-площадь-под-кривой-ошибок/>

Подумайте, почему мы вводили ROC-AUC **не** с помощью движения по сетке вправо и вверх (подсказка: ответ кроется в шаге с сортировкой)

3. Выбор метрики (пример: рекомендации)

Что можем делать

- Прогнозировать, какие товары будут куплены
- Максимизировать прибыль

Остается вопрос: какие прогнозы нужны и как их использовать, чтобы денег стало больше?

Максимизация количества покупок

Товар 1	Товар 2	Товар 3	Товар 4
---------	---------	---------	---------

Максимизация количества покупок

Товар 1	Товар 2	Товар 3	Товар 4
---------	---------	---------	---------

Вероятность:

p_1

p_2

p_3

p_4

Максимизация дохода

Товар 1	Товар 2	Товар 3	Товар 4
---------	---------	---------	---------

Вероятность:	p_1	p_2	p_3	p_4
Цена:	c_1	c_2	c_3	c_4

Максимизация дохода



Puma
Ветровка
3 490 руб.



Crocs
Сланцы
1 990 руб.



Tony-p
Слипоны
~~1 999 руб.~~ 1 590 руб.



Champion
Брюки спортивные
~~3 599 руб.~~ 1 970 руб.

Вероятность:	0.05	0.02	0.015	0.009
Цена:	3490	1990	1590	1970

Максимизация прибыли



Puma
Ветровка
3 490 руб.



Crocs
Сланцы
1 990 руб.



Tony-p
Слипоны
~~1 999 руб.~~ 1 590 руб.



Champion
Брюки спортивные
~~3 599 руб.~~ 1 970 руб.

Вероятность:	0.05	0.02	0.015	0.009
Цена:	3490	1990	1590	1970
Маржинальность	0.1	0.4	0.4	0.2

Мини-задача

Как изменится построение модели, если нам нужно максимизировать количество просмотренных пользователем товаров?

Точность (Precision@k)

Рекомендованные товары
Синяя футболка
Красная футболка
Кроссовки
Кепка
Зеленая футболка

Купленные товары
Красная футболка
Кеды
Кепка

k — количество
рекомендаций

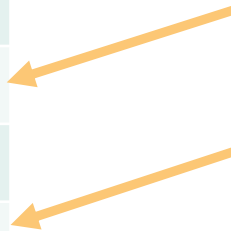
$$\text{Precision@}k = \frac{\text{купленное из рекомендованного}}{k}$$

AveragePrecision@k - усредненный по сессиям Precision@k

Полнота (Recall@k)

Рекомендованные товары
Синяя футболка
Красная футболка
Кроссовки
Кепка
Зеленая футболка

Купленные товары
Красная футболка
Кеды
Кепка



k — количество
рекомендаций

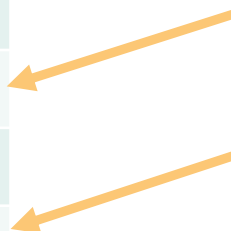
$$\text{Recall@k} = \frac{\text{купленное из рекомендованного}}{\text{количество покупок}}$$

AverageRecall@k - усредненный по сессиям Recall@k

Взвешенный ценами recall@k

Рекомендованные товары
Синяя футболка – 1000р
Красная футболка – 1200р
Кроссовки – 3500р
Кепка – 900р
Зеленая футболка – 800р

Купленные товары
Красная футболка – 1200р
Кеды – 3000р
Кепка – 900р



$$\text{Взвешенный ценами Recall@k} = \frac{\text{стоимость купленного из рекомендованного}}{\text{стоимость покупок}}$$

AverageRecall@k - усредненный по сессиям Recall@k

Качество классификации против качества рекомендаций

Пример – 2 решения для прогноза купит/не купит товар:

	Алгоритм 1	Алгоритм 2
AUC классификатора	0.52	0.85
Recall@5	0.72	0.71

План

1. Метрики в задачах регрессии

2. Метрики в задачах классификации

3. Пример выбора метрики

Data Mining in Action

Лекция 5

Группа курса в Telegram:



<https://t.me/joinchat/B1OlTk74nRV56Dp1TDJGNA>