



Метрики оценки качества



План

1. Регрессия

2. Классификация

3. Примеры

+ Бонус

1. Регрессия



Метрики качества

RMSE

MAE

R^2

MAPE

SMAPE

ROOT MEAN SQUARED ERROR

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

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$
$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

MEAN AVERAGE ERROR

- Отклонение прогноза от исходного значения
- Усредненное по всем наблюдениям

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

R^2 – COEFFICIENT OF DETERMINATION

- $$R^2 = 1 - \frac{MSE(model)}{MSE(const)}$$

Показывает долю информации (дисперсии), которую «объясняет» модель

MEAN AVERAGE PERCENTAGE ERROR

- Ошибка прогнозирования оценивается в процентах

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

SYMMETRIC MEAN AVERAGE PERCENTAGE ERROR

- Ошибка прогнозирования 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 прогнозирования оценивается в процентах

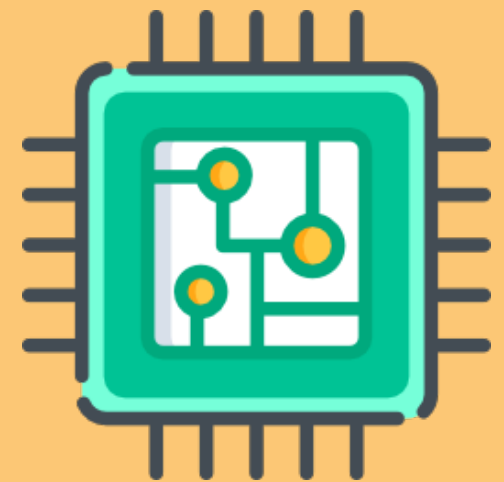
$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

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

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

2. Классификации



Метрики

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

Precision & Recall

- Precision – точность
- Recall – полнота

Accuracy & baselines

α — доля 1 в выборке

β — вероятность 1 в рандомном предсказании

Вероятность угадать

$$\alpha\beta + (1 - \alpha)(1 - \beta) = 1 - \alpha + \beta(2\alpha - 1)$$

При несбалансированной выборке имеем разные скоры

Сбитые самолеты



Сбитые самолеты



$$y = (0000101101)$$

$$\hat{y} = (0110100101) \leftarrow$$



Precision

Precision – точность выстрелов

$$\frac{\text{Количество сбитых самолётов}}{\text{Количество выстрелов}}$$

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

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



Recall

Recall – «полнота» сбивания

Количество сбитых самолётов

Общее количество самолетов

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

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



Confusion matrix

		True Class	
		Positive (P)	Negative (N)
Predicted Class	Positive (+)	True Positive Count (TP)	False Positive Count (FP)
	Negative (-)	False Negative Count (FN)	True Negative Count (TN)

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

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

Precision & recall & baselines

α — доля 1 в выборке

β — вероятность 1 в рандомном предсказании

$$Precision = \frac{\alpha\beta}{\beta} = \alpha \qquad Recall = \frac{\alpha\beta}{\alpha} = \beta$$

Precision не зависит от баланса предсказаний

Recall не зависит от баланса классов

F-measure (F-score, F1)

- Среднее гармоническое между precision и recall
- Значение F-measure ближе к меньшему из precision, recall

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

ROC-AUC

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

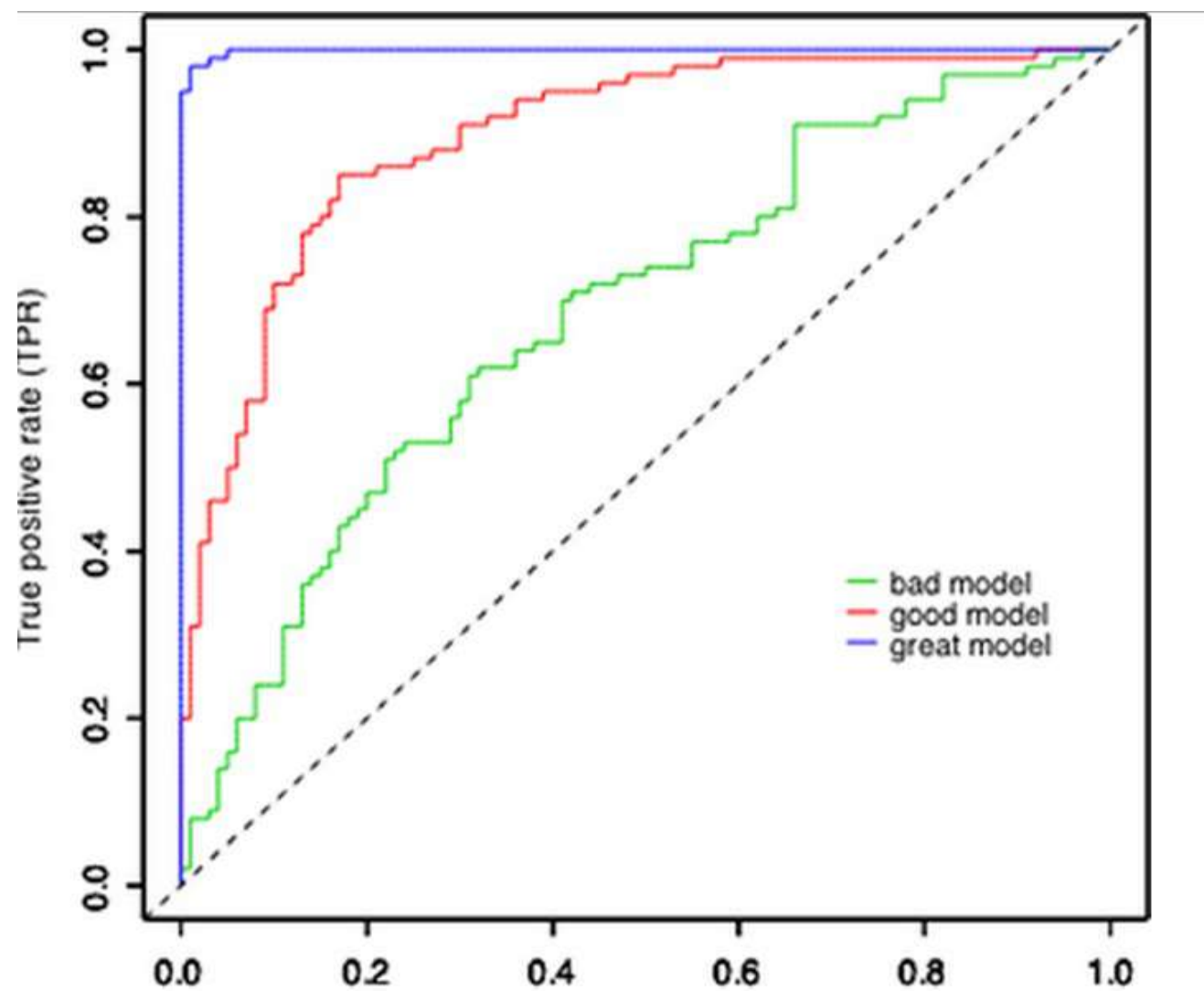
ROC

		True Class	
		Positive (P)	Negative (N)
Predicted Class	Positive (+)	True Positive Count (TP)	False Positive Count (FP)
	Negative (-)	False Negative Count (FN)	True Negative Count (TN)

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

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

ROC



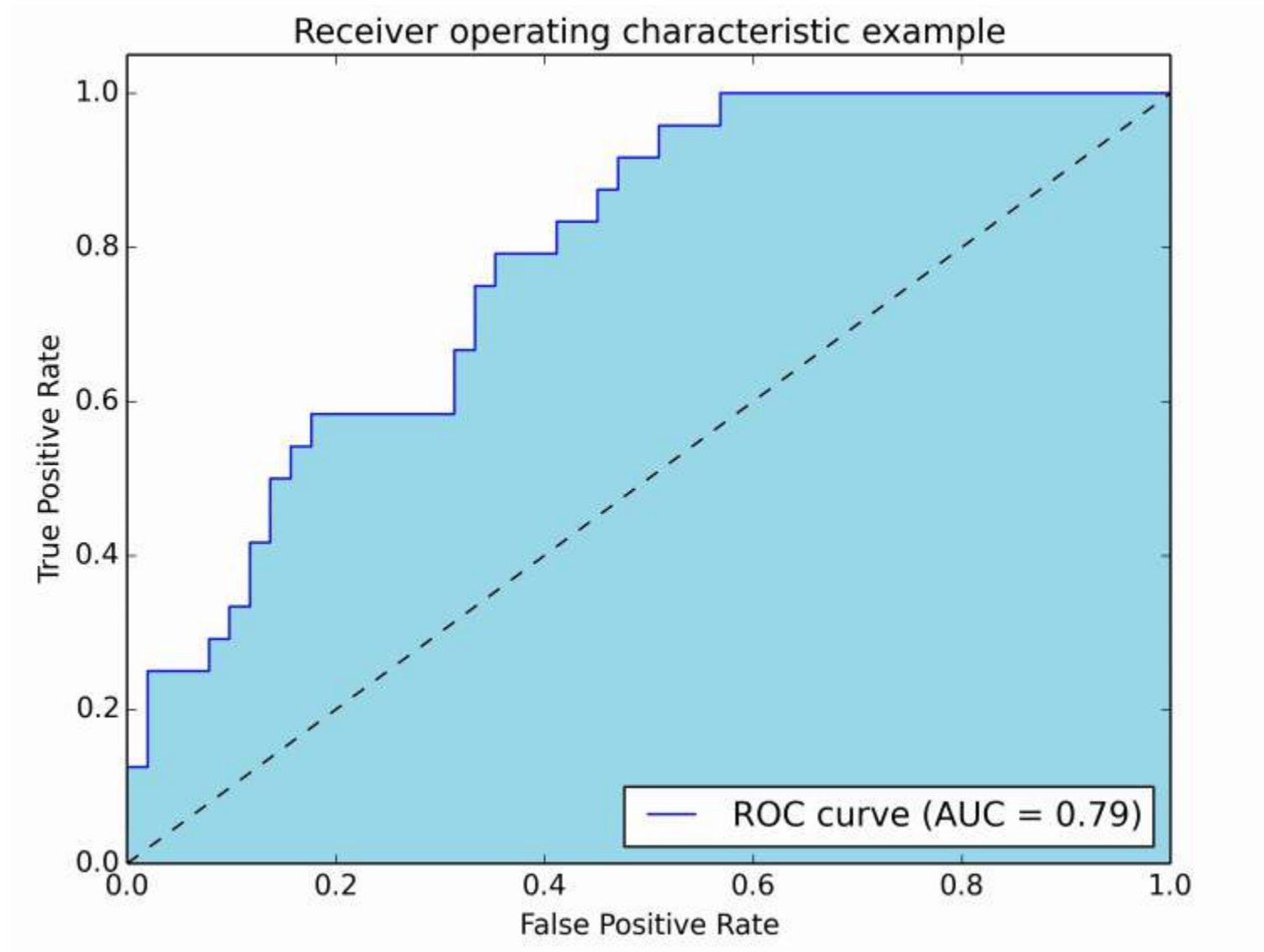
ROC-AUC

- Как оценить кривую численно?

ROC-AUC

- Как оценить кривую численно?
- Измерить площадь под кривой – Area Under Curve

ROC-AUC



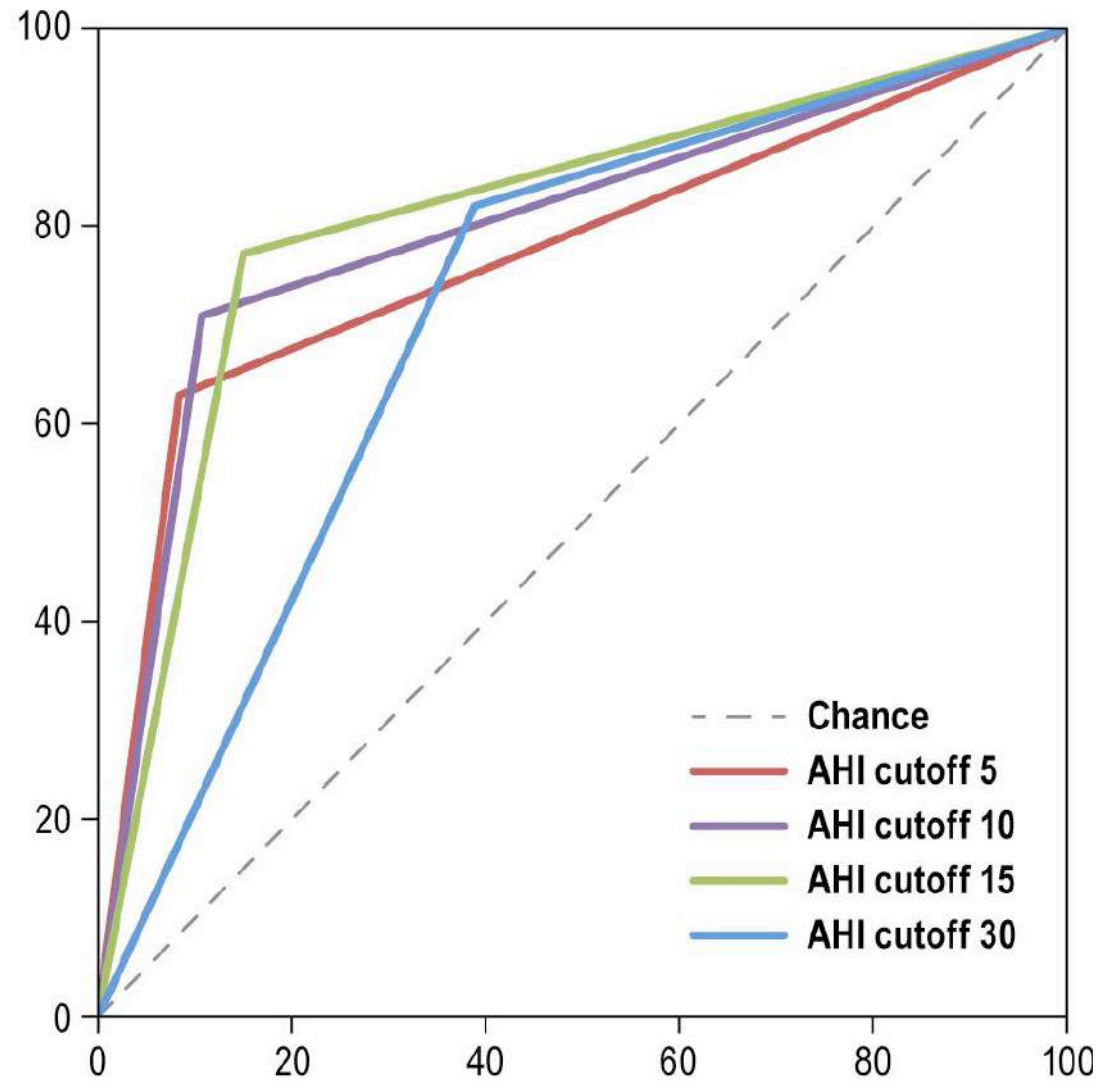
ROC-AUC

- Что если классификация всё же не вероятностная?

ROC-AUC

- Что если классификация всё же не вероятностная?
- Строим кривую по **трем** точкам

ROC-AUC



3. Примеры



Выбираем, что оптимизировать (на примере рекомендаций)

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

Автоматизация процесса

Хотим наилучшее качество, есть ограничение на количество обработанных объектов

Хотим оптимизировать costs, есть ограничение на качество обработки

Автоматизация процесса

Хотим наилучшее качество, есть ограничение на количество обработанных объектов

max precision on fixed recall

Хотим оптимизировать касты, есть ограничение на качество обработки

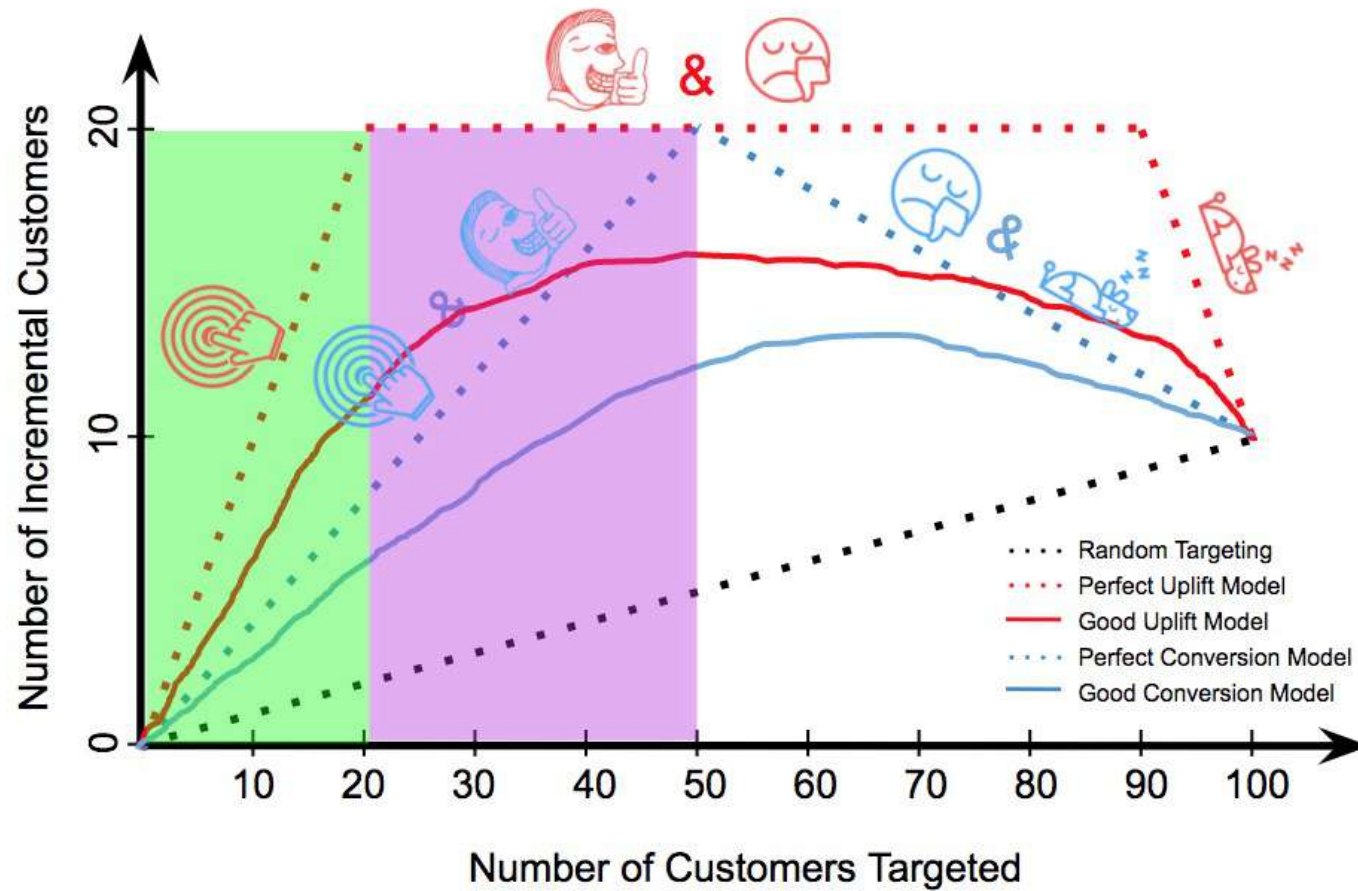
max recall on fixed precision

Uplift

Как оценить качество в задачах таргетирования?

Например,
звонка для реактивации пользователя в сервисе

Uplift



Бонус



Как оптимизировать метрику качества

- Напрямую: MSE, MAE
- Напрямую с извращениями: MAPE, Uplift, blackbox
- Перебором гиперпараметров: precision, recall, accuracy

MAPE

Способ 1

- $w_i = 1/|y_i|$

Способ 2

- $y_i^{new} = \ln y_i$ & use MAE
- $|\ln y - \ln a| = \left| \ln \left(1 - \frac{a-y}{y} \right) \right| \approx \left| \frac{y-a}{y} \right|$

Uplift

Способ 1

- $Model(x)$ – средний выигрыш на объекте x
- $uplift = model_{exp}(x) - model_{control}(x)$

Способ 2

- $Model(x, exp)$ – средний выигрыш на объекте x в группе exp
- $uplift = model(x, exp = 1) - model(x, exp = 0)$

Uplift

Способ 3

- Target – выигрыш на объекте $x * (2 I\{exp = 1\} - 1)$
- $Model(x)$ = средний таргет на объекте x
- $uplift = model(x)$

Blackbox

- **Random Search, Grid Search**
- **Simulated annealing, Differential Evolution**
- **Bayesian Optimization**

Как оптимизировать метрику качества

- Напрямую: MSE, MAE
- Напрямую с извращениями: MAPE, Uplift, blackbox
- Перебором гиперпараметров: precision, recall, accuracy