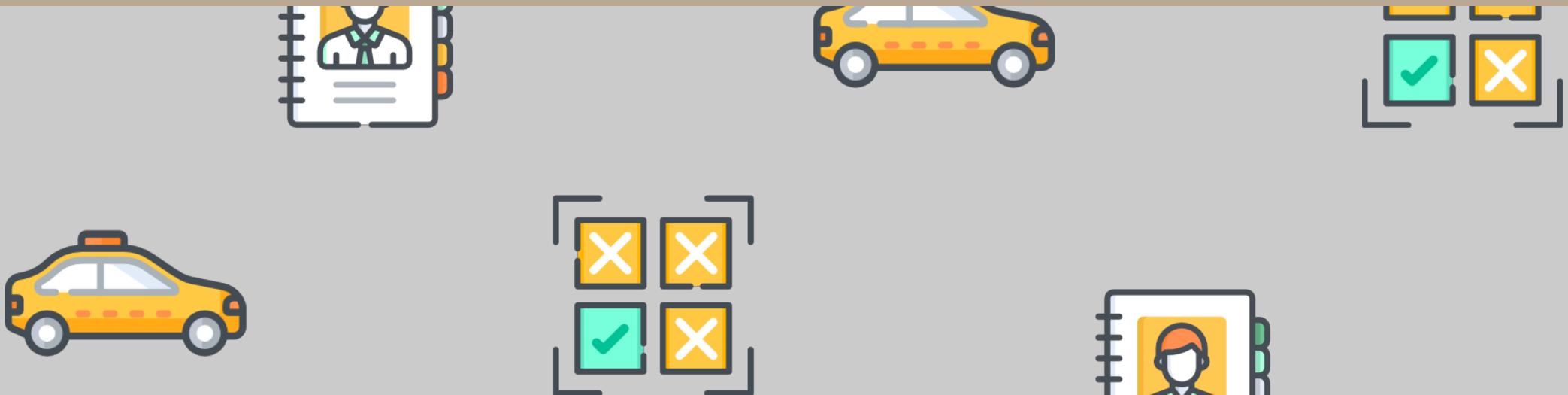


Метрики оценки качества



На этой лекции

- I. Метрики качества в задачах регрессии
- II. Метрики качества в задачах классификации
- III. Выбор метрик качества: пример
- IV. Устойчивость моделей

I. Метрики качества в задачах регрессии

Метрики качества

- 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:

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

LogLoss

- Логарифмическая ошибка
- Хорошо оценивает вероятность

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

II. Метрики качества 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 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 1 0 0

predicted: 0 0 1 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

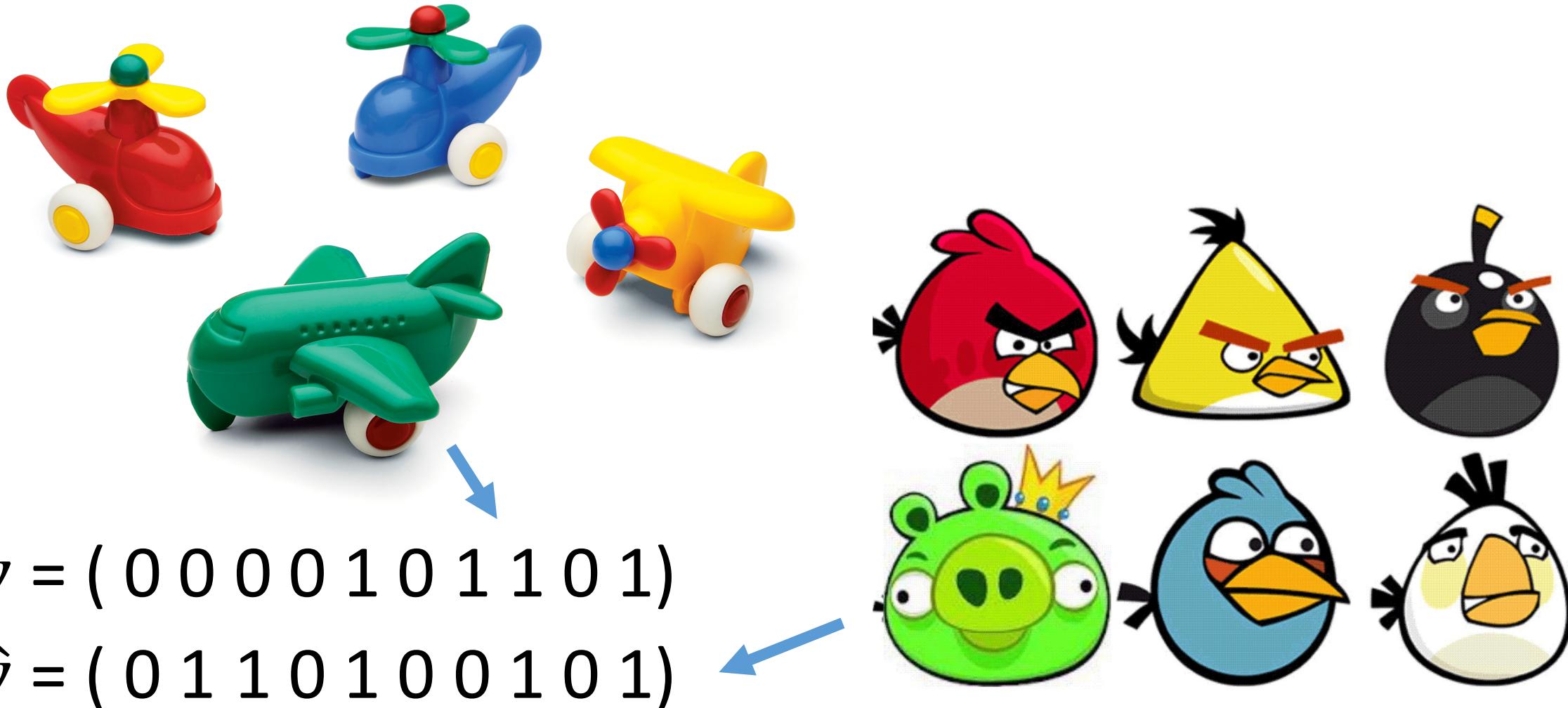
Precision & Recall

- Precision – точность
- Recall - полнота

Сбитые самолеты



Сбитые самолеты



Precision

Precision – точность выстрелов:

Количество сбитых самолётов

Количество выстрелов

$$y = (0\ 0\ 0\ 0\ \textcolor{red}{1}\ 0\ 1\ \textcolor{red}{1}\ 0\ 1)$$

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



Recall

Recall – «полнота» сбивания самолетов:

Количество сбитых самолётов

Общее количество самолётов

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

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



Связь с True Positive, False Positive и др.

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

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

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

F-measure (F-score, F1)

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

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

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.

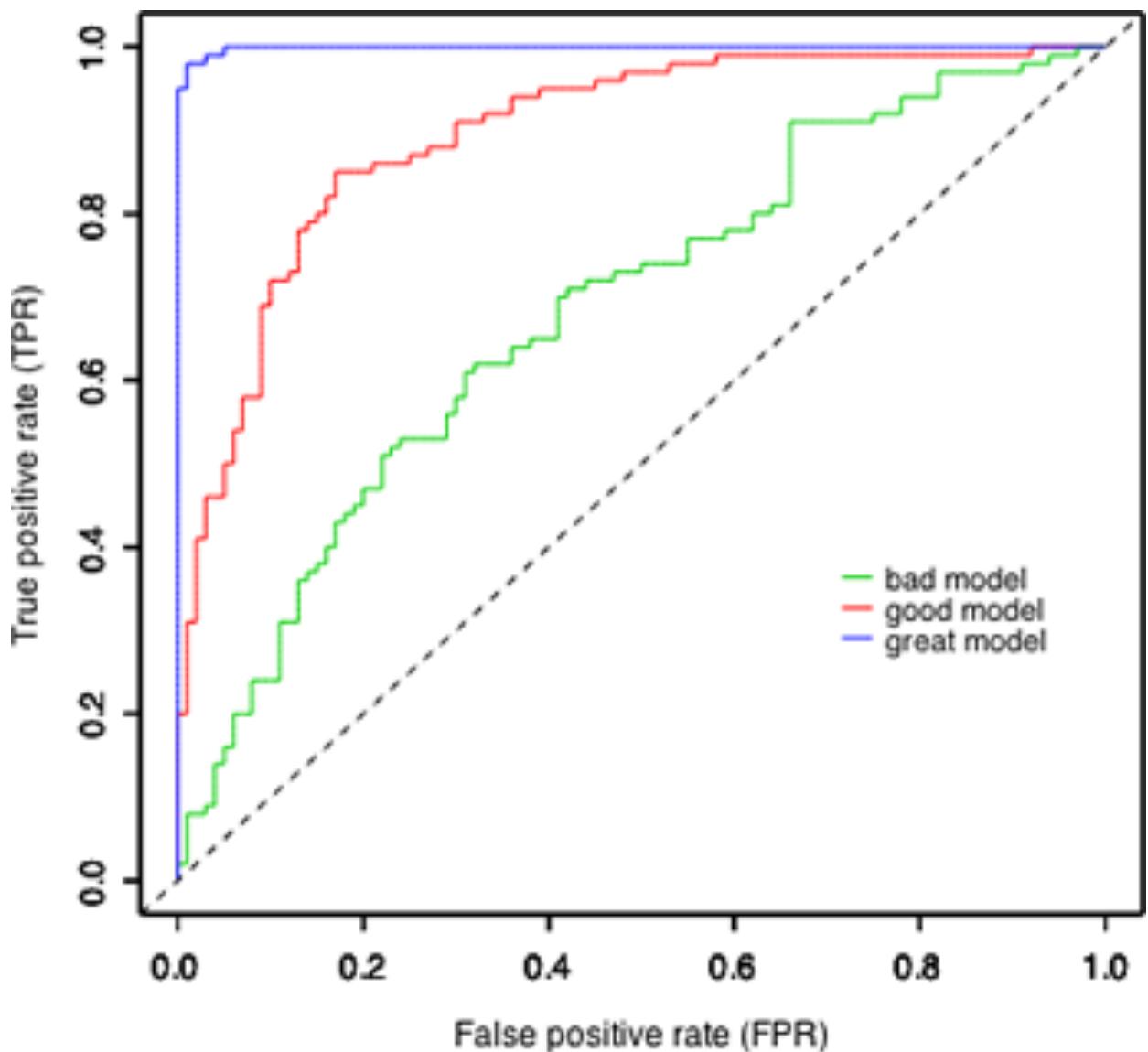
ROC

		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



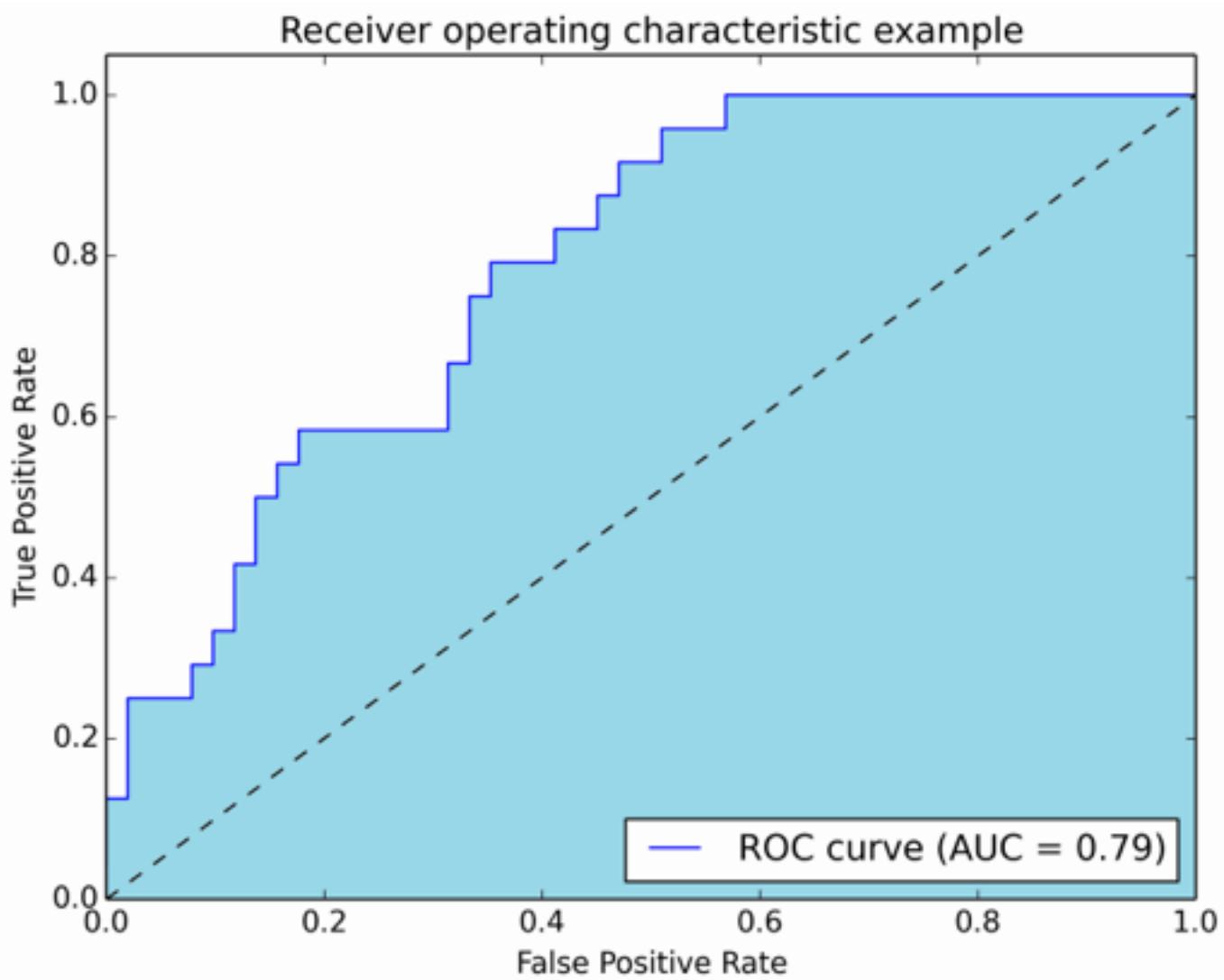
ROC

- Как оценить кривую численно?

ROC-AUC

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

ROC-AUC



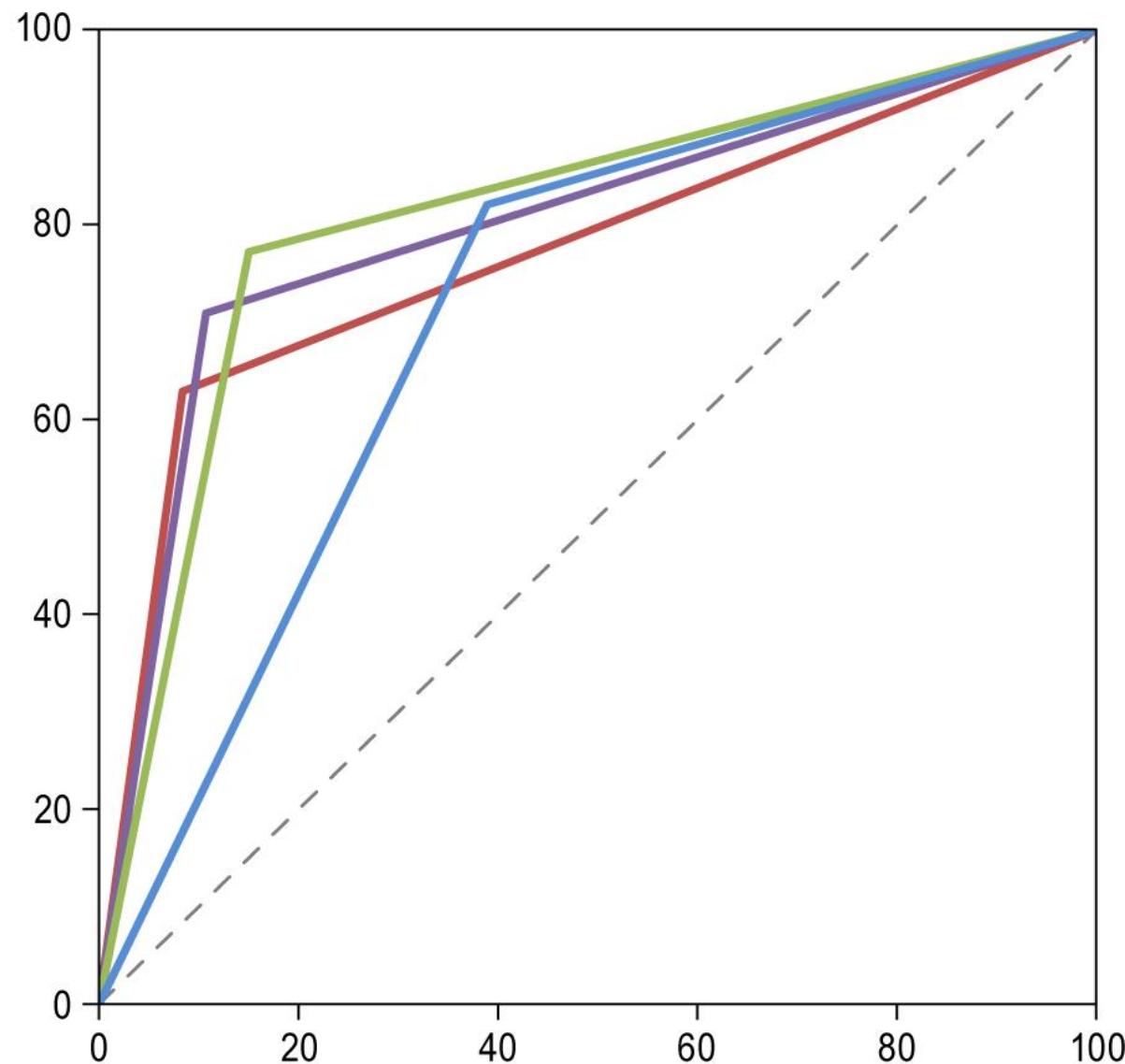
ROC-AUC

- Что если классификация всё же не вероятностная?

ROC-AUC

- Что если классификация всё же не вероятностная?
- Строим кривую по 3м точкам

ROC-AUC



III. Выбор метрик качества: пример

1. 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

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?