

Метрики и валидация

Лекция 2

Разминочный тест на оценку

<https://forms.gle/J11gVygAK2UDYmM69>



Повторение

X – множество объектов (их признаков описание)

Y – множество истинных ответов

\hat{Y} – множество предсказанных ответов. Получаем по формуле:

$$\hat{Y} = f(X),$$

где $f(X)$ -- модель машинного обучения

Чем ближе \hat{Y} к Y , тем лучше аппроксимируем неизвестную зависимость => лучше модель

На чем оценивать близость \hat{Y} и Y

Разбиваем датасет на два множества – **обучающее** и **тестовое**

$$X = X_{train} \cup X_{test}$$

$$Y = Y_{train} \cup Y_{test}$$

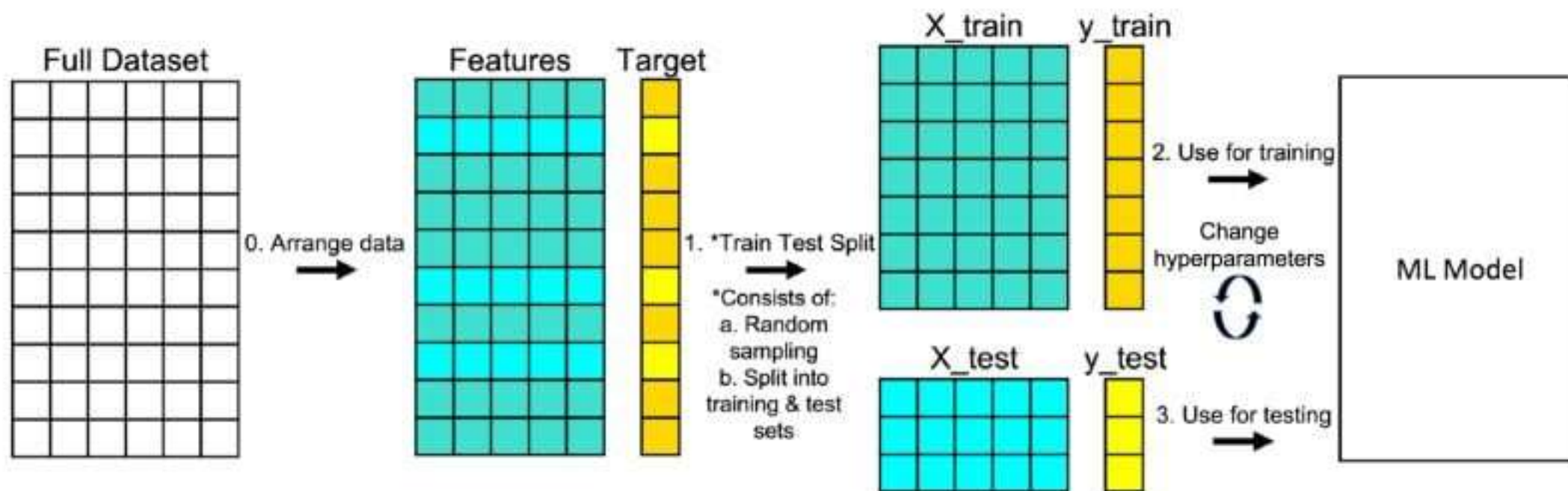
$$\hat{Y} = \hat{Y}_{train} \cup \hat{Y}_{test}$$

Dataset

Training

Testing

Разбиение на обучение и тест



Алгоритм построения и валидации модели

1. Разбиение датасета на X, Y
2. Разбиение множеств X, Y на обучающее и тестовое множества
3. Обучение модели $f()$ на (X_{train}, Y_{train})
4. Получение предсказаний $\hat{Y}_{test} = f(X_{test})$ и $\hat{Y}_{train} = f(X_{train})$
5. Оцениваем близость \hat{Y}_{test} и Y_{test} , а также Y_{train} и \hat{Y}_{train} .

Близость множеств оцениваются при помощи **метрик**.

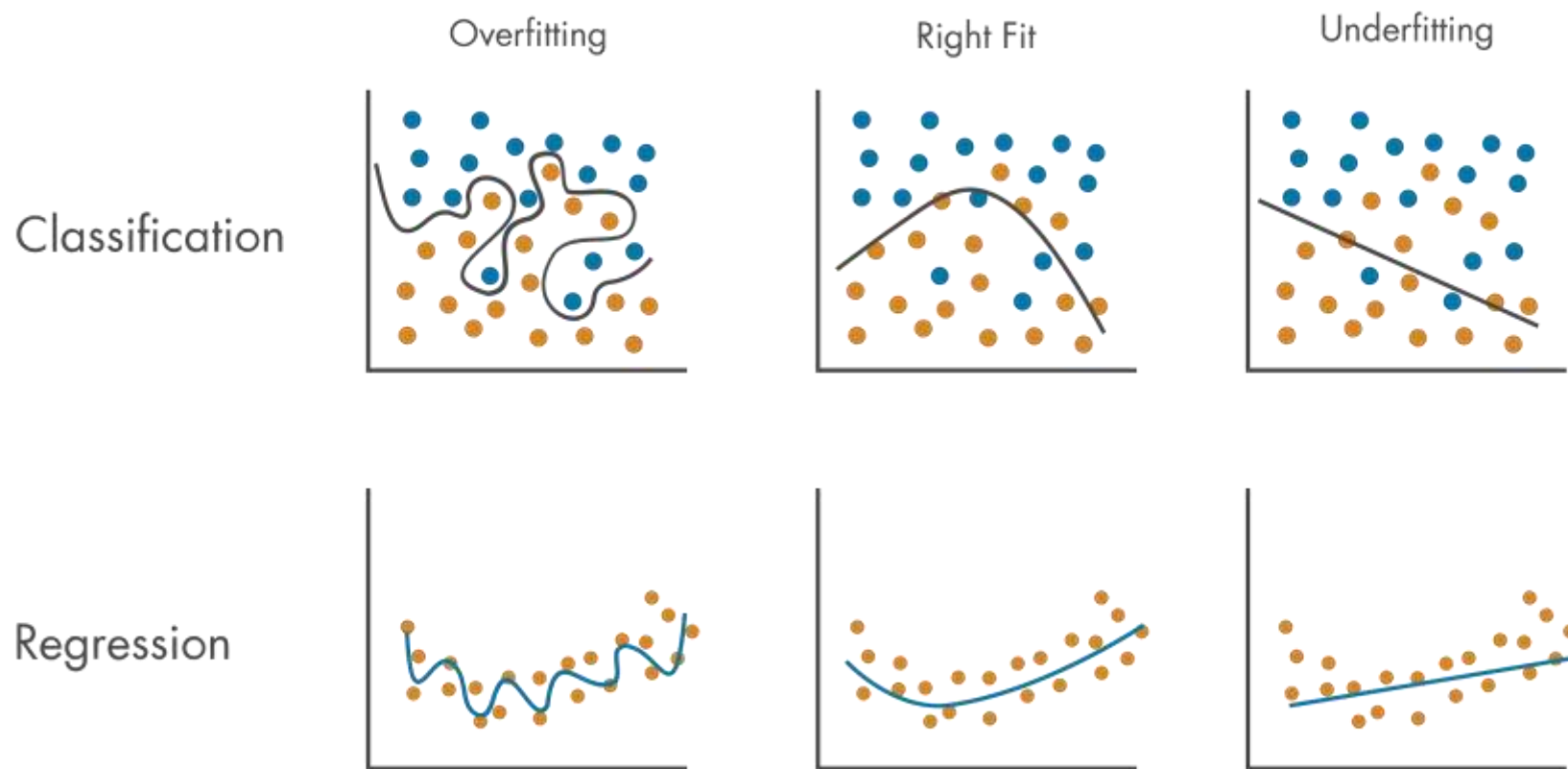
Разбиение на тестовое и обучающее множество необходимо, чтобы избежать **переобучения**.

Переобучение и метрики

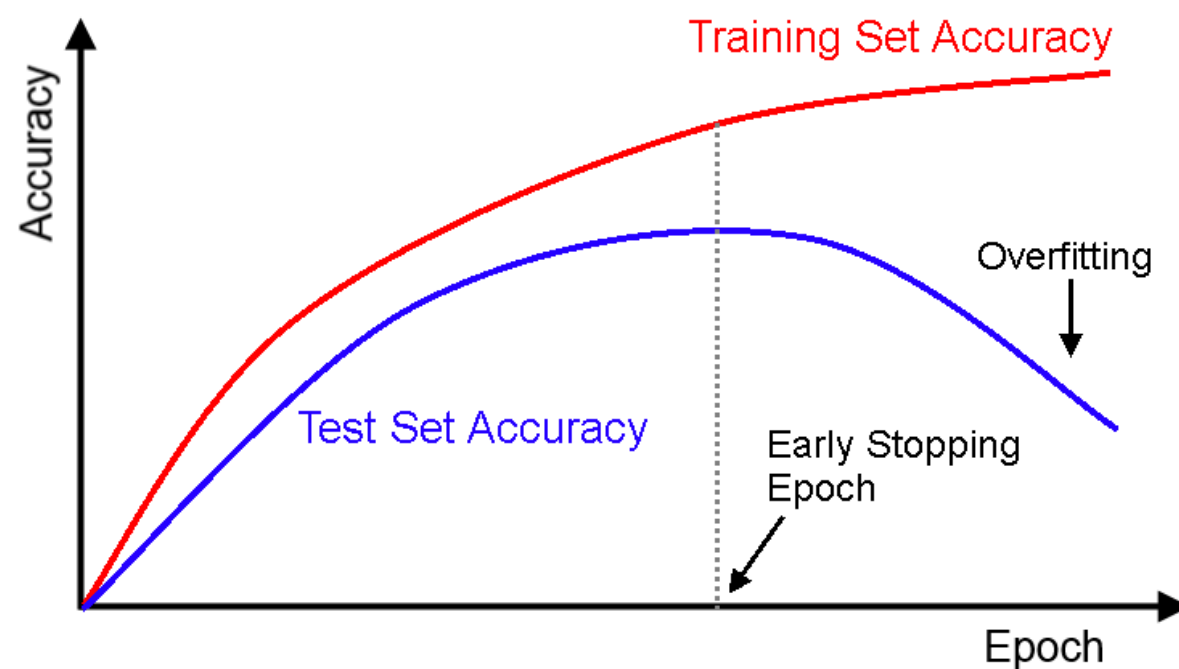
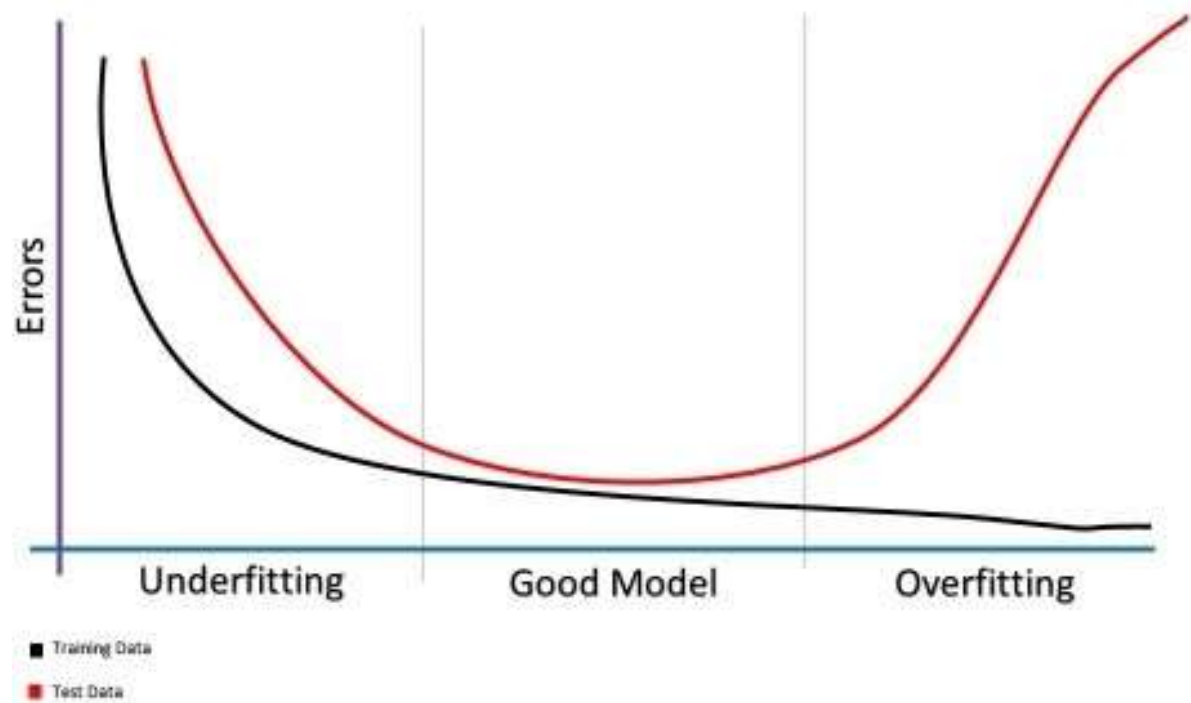
Переобучение (*переподгонка*, *пере-* в значении «слишком», ***overfitting***) — явление, когда построенная модель хорошо объясняет примеры из **обучающей выборки**, но относительно плохо работает на примерах, не участвовавших в обучении (на примерах из **тестовой выборки**).

Метрика — критерий качества модели машинного обучения.

Переобучение графически



Переобучение на лоссах / метриках



Метрики. Классификация

1. Accuracy – доля правильно предсказанных классов

$$Accuracy = \frac{\textit{correct class}}{\textit{total class}} = \frac{1}{N} \sum_{i=1}^N [y_i = \hat{y}_i]$$

$$Accuracy \in [0, 1]$$

$$Accuracy \in [0\%, 100\%]$$

| | true | predict |
|---|------|---------|
| 0 | 0 | 0 |
| 1 | 0 | 1 |
| 2 | 0 | 0 |
| 3 | 0 | 0 |
| 4 | 0 | 0 |
| 5 | 0 | 0 |
| 6 | 1 | 1 |
| 7 | 1 | 1 |
| 8 | 0 | 0 |
| 9 | 0 | 0 |

Метрики. Классификация. Дисбаланс

Множество Y :

0 – 990 объектов

1 – 10 объектов

Предсказываем **всегда** самый популярный класс:
 $f(X) = 0$

В таком случае $accuracy = 99\%$

Метрики. Матрица ошибок

Теперь есть целых 4 варианта, куда может попасть пара (y_i, \hat{y}_i) :

TP, FP, FN, TN

Можно переписать формулу для accuracy:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Confusion Matrix

| | Actually Positive (1) | Actually Negative (0) |
|------------------------|-----------------------|-----------------------|
| Predicted Positive (1) | True Positives (TPs) | False Positives (FPs) |
| Predicted Negative (0) | False Negatives (FNs) | True Negatives (TNs) |

Метрики. Precision

Precision – точность

$$\text{Precision} = \frac{TP}{TP + FP}$$

По-человечески: доля истинно целевых событий среди всех событий, отмеченных моделью как целевое.

Примеры:

1. Детектирование редкого заболевания через тест-систему: положительный результат у 100 человек, из них действительно болеют 95. Тогда *Precision* = 95%.
2. Мошеннические операции: заблокировали 1000 транзакций, действительно мошеннических из них 200. *Precision* = 20%.

Confusion Matrix

| | Actually Positive (1) | Actually Negative (0) |
|------------------------|-----------------------|-----------------------|
| Predicted Positive (1) | True Positives (TPs) | False Positives (FPs) |
| Predicted Negative (0) | False Negatives (FNs) | True Negatives (TNs) |

Метрики. Recall

Recall – полнота

$$\text{Recall} = \frac{TP}{TP + FN}$$

По-человечески: доля целевых событий, отловленных моделью, относительно всех истинных целевых событий.

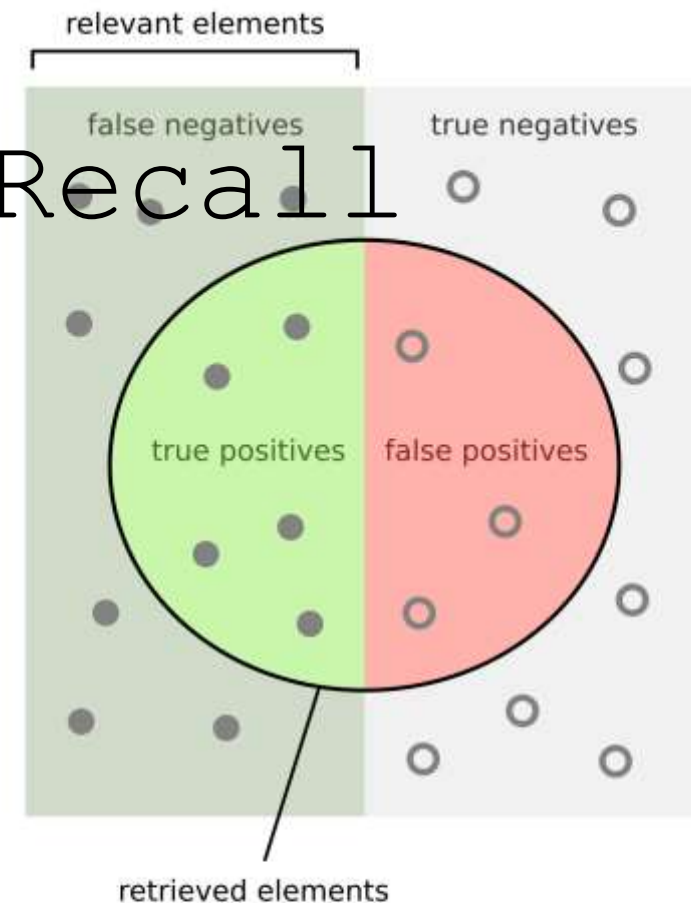
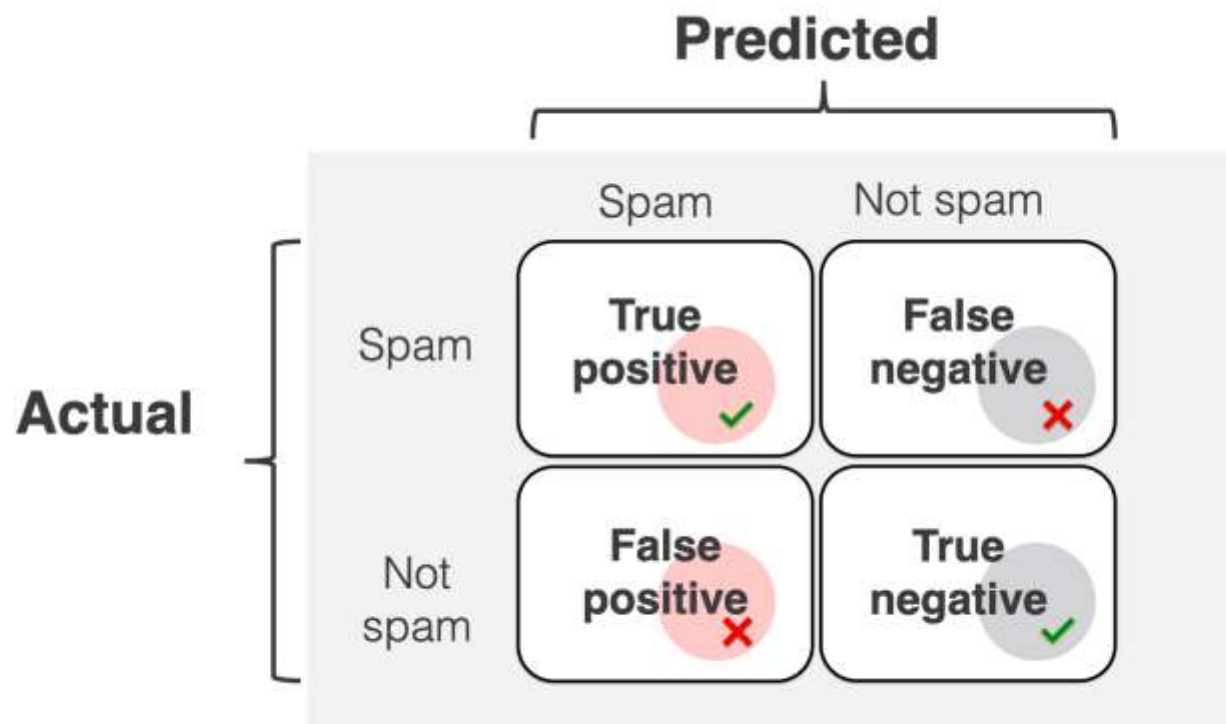
Примеры:

1. Детектирование редкого заболевания через тест-систему: положительный результат у 100 человек, из них действительно болеют 95. Всего болеющих 1000. Тогда $\text{Recall} = 9.5\%$.
2. Мошеннические операции: заблокировали 1000 транзакций, действительно мошеннических из них 200. Всего было совершено 250 мошеннических транзакций. $\text{Recall} = 80\%$.

Confusion Matrix

| | Actually Positive (1) | Actually Negative (0) |
|------------------------|-----------------------|-----------------------|
| Predicted Positive (1) | True Positives (TPs) | False Positives (FPs) |
| Predicted Negative (0) | False Negatives (FNs) | True Negatives (TNs) |

Метрики. Precision и Recall



How many retrieved items are relevant?

Precision =



How many relevant items are retrieved?

Recall =

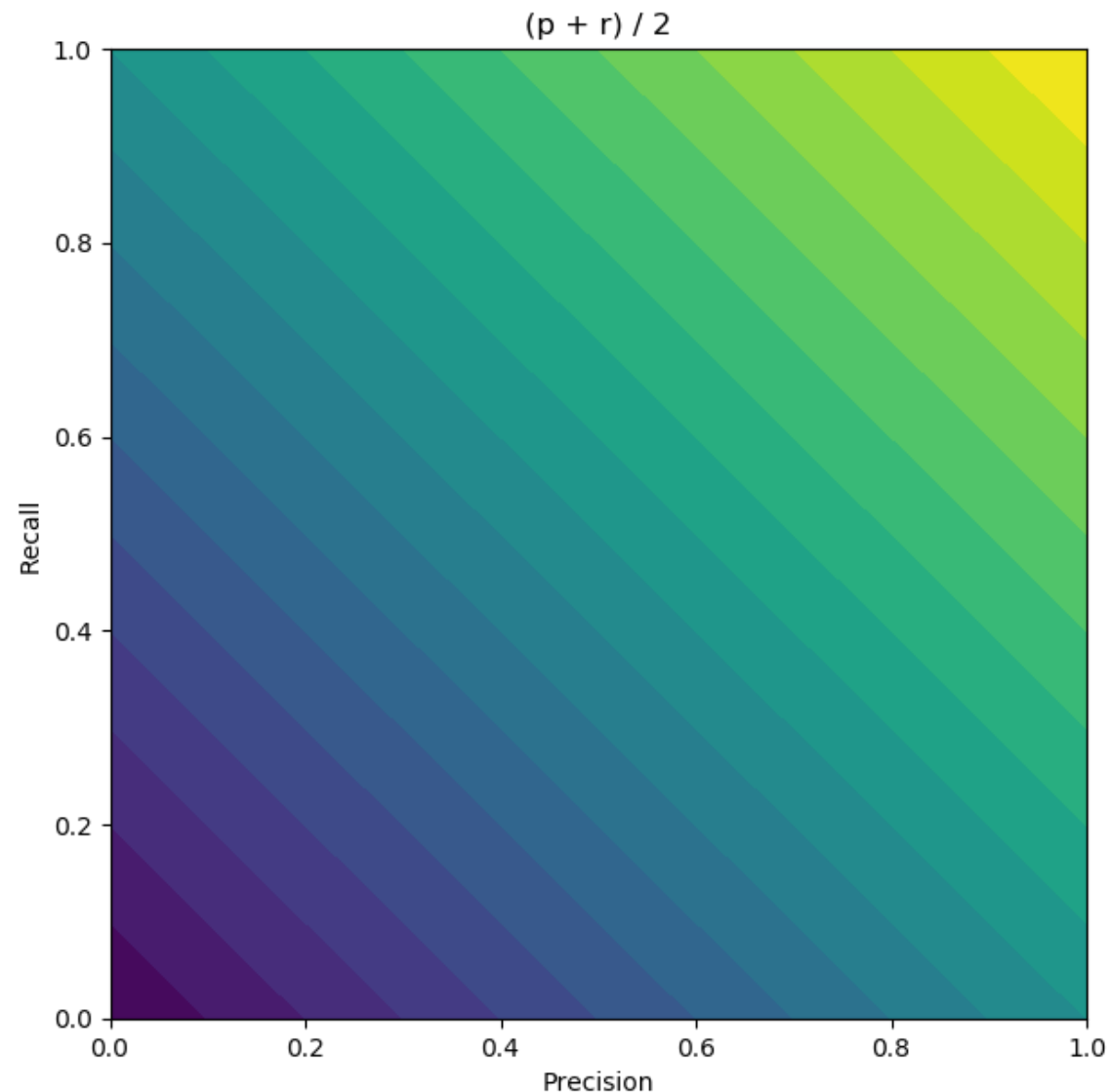


Метрики. F-мера

Как-нибудь учитывать
одновременно *Recall* и
Precision

Например:

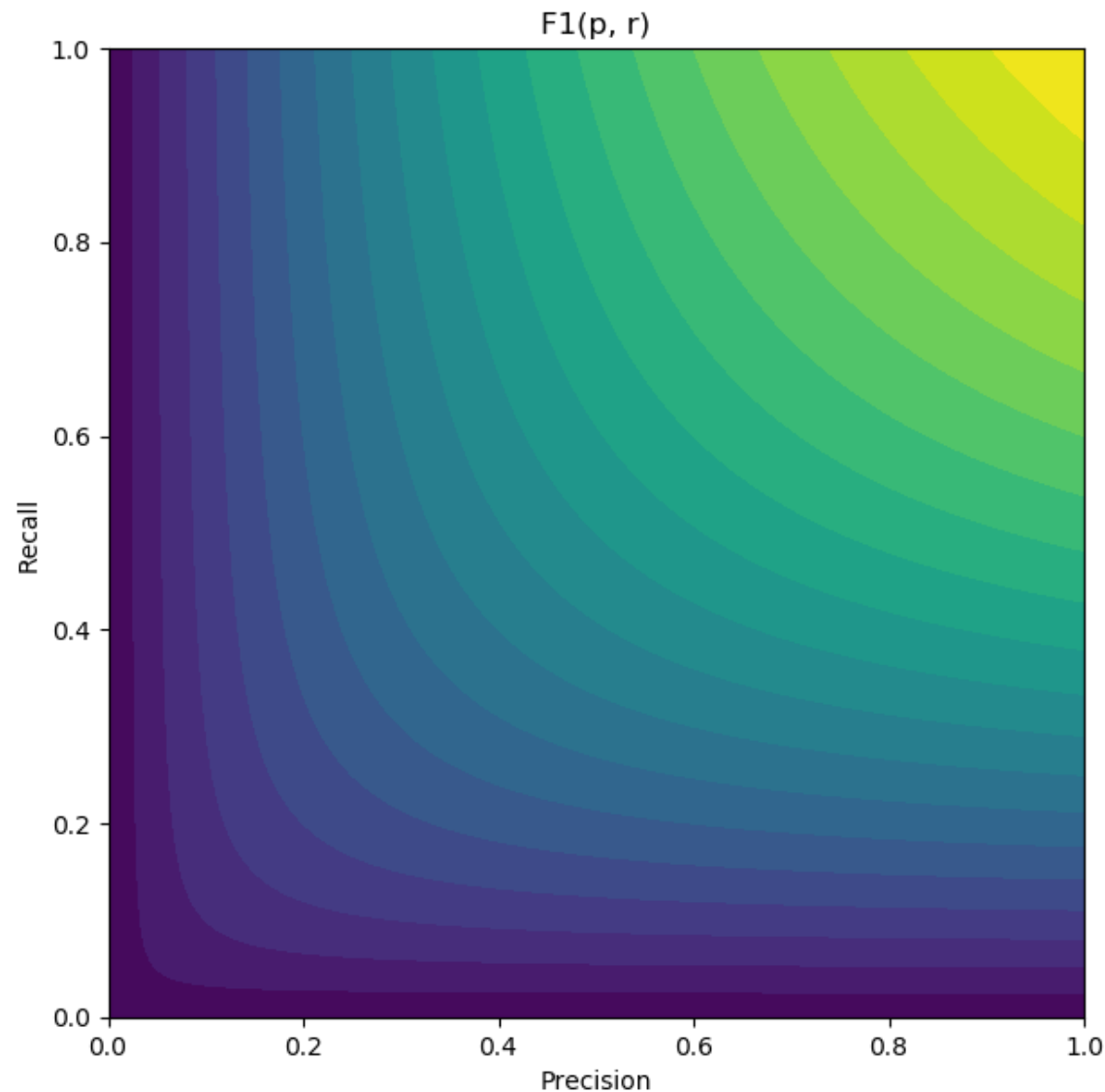
$$\frac{Recall + Precision}{2}$$



Метрики. F-мера

Лучше делать вот так:

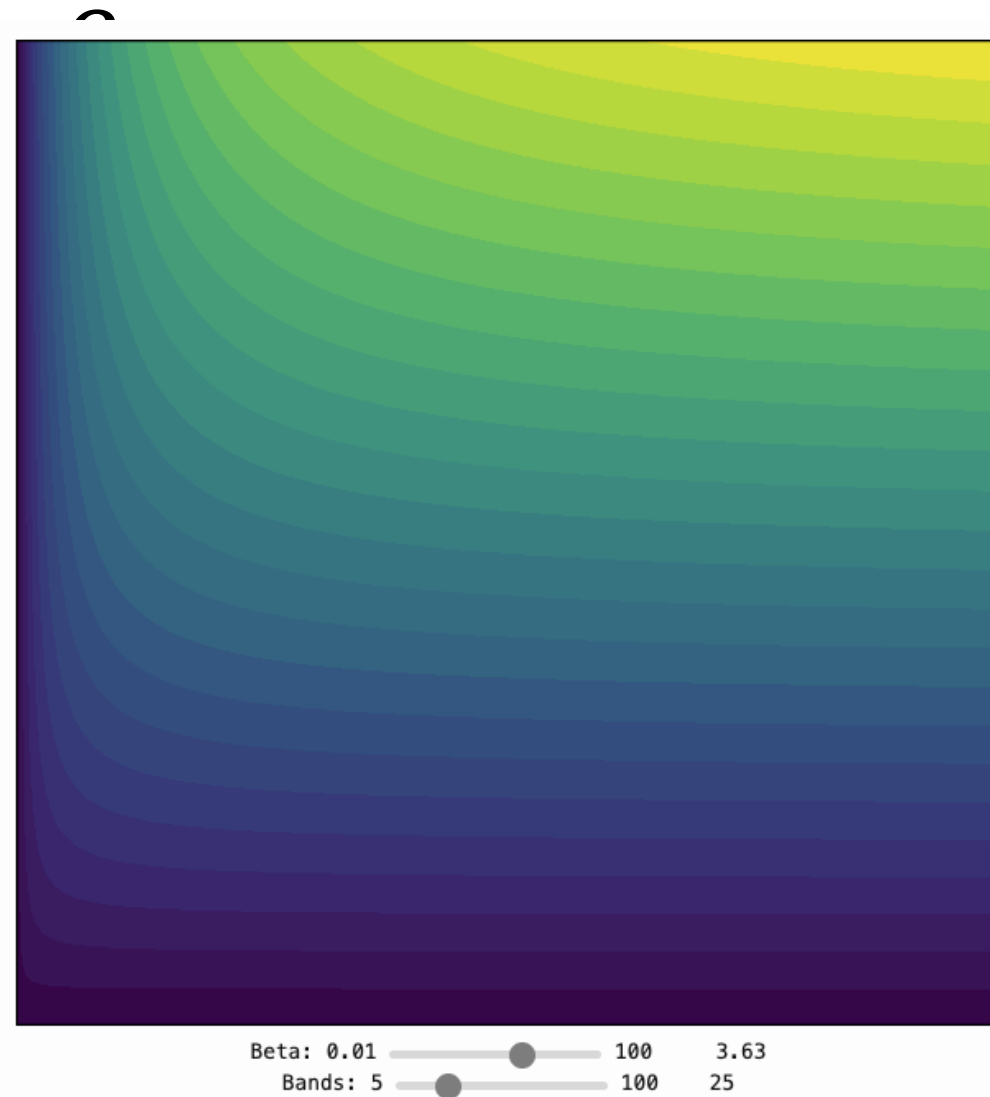
$$F = 2 \frac{Precision \cdot Recall}{Precision + Recall}$$



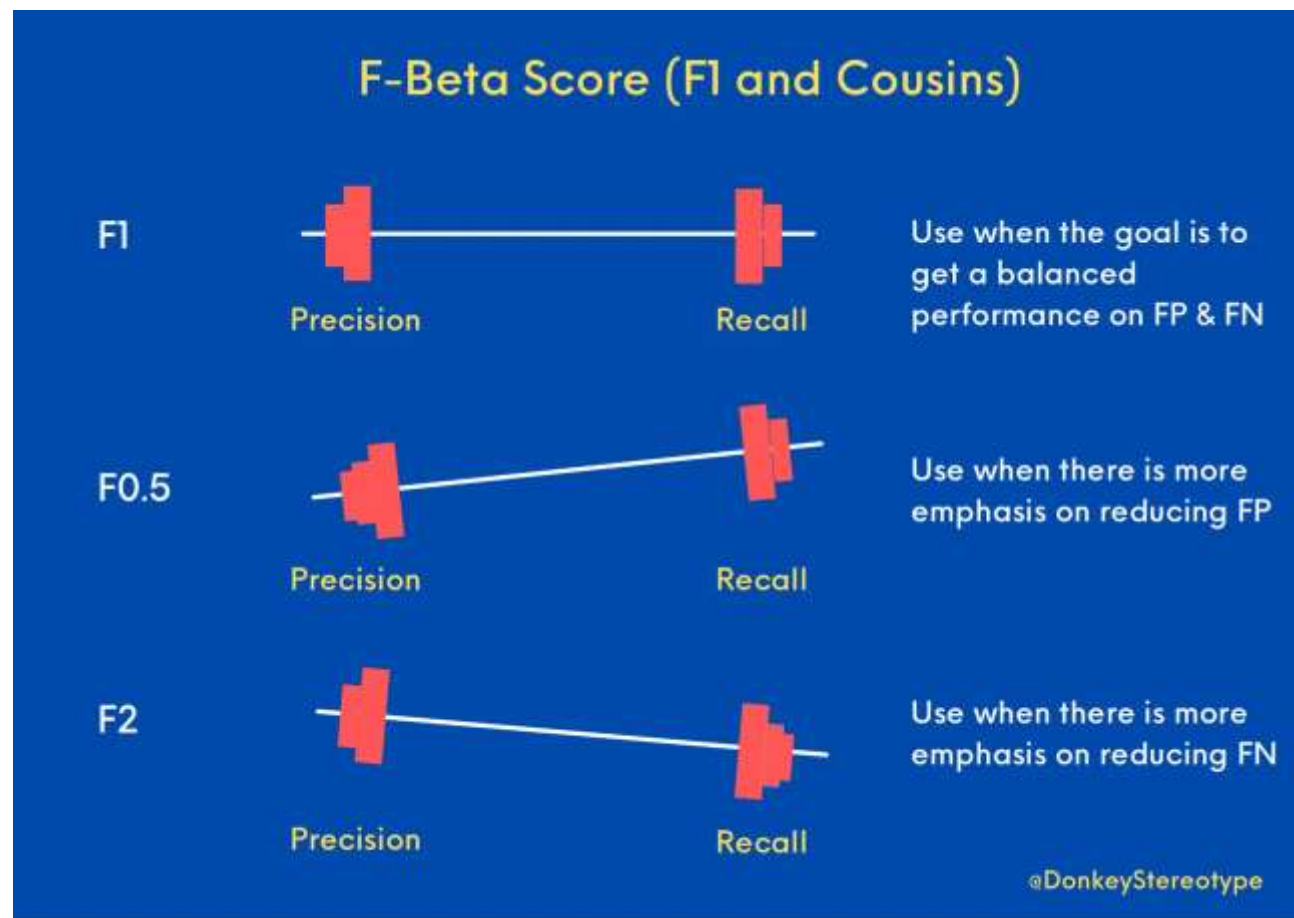
Метрики. F-мера с

Можно учитывать разные
цены ошибок с помощью
параметра β :

$$F_{\beta} = (1 + \beta^2) \frac{Precision \cdot Recall}{\beta^2 \cdot Precision + Recall}$$



F-мера с β



Метрики. ROC-AUC

Метрика, отражающая
ранжирующую способность модели
Метрика не зависит от порога
Устойчива к дисбалансу

| | true | predict |
|-------------|------|----------|
| 0 | 0 | 0.000012 |
| 1 | 0 | 0.000218 |
| 2 | 0 | 0.008439 |
| 3 | 0 | 0.002587 |
| 4 | 0 | 0.000249 |
| ... | ... | ... |
| 2995 | 0 | 0.011185 |
| 2996 | 0 | 0.206901 |
| 2997 | 0 | 0.105899 |
| 2998 | 0 | 0.000023 |
| 2999 | 0 | 0.000145 |

Метрики. ROC-AUC

Строится график в координатах FPR, TPR:

$$FPR = \frac{FP}{TN + FP}$$

$$TPR = \frac{TP}{TP + FN}$$

Считаются пары (FPR, TPR) на всех порогах и наносятся на график

| | true | predict |
|-------------|------|----------|
| 0 | 0 | 0.000012 |
| 1 | 0 | 0.000218 |
| 2 | 0 | 0.008439 |
| 3 | 0 | 0.002587 |
| 4 | 0 | 0.000249 |
| ... | ... | ... |
| 2995 | 0 | 0.011185 |
| 2996 | 0 | 0.206901 |
| 2997 | 0 | 0.105899 |
| 2998 | 0 | 0.000023 |
| 2999 | 0 | 0.000145 |

Метрики. ROC-AUC. Пример

| id | оценка | класс |
|----|--------|-------|
| 1 | 0.5 | 0 |
| 2 | 0.1 | 0 |
| 3 | 0.2 | 0 |
| 4 | 0.6 | 1 |
| 5 | 0.2 | 1 |
| 6 | 0.3 | 1 |
| 7 | 0.0 | 0 |

Табл. 1

| id | оценка | класс |
|----|--------|-------|
| 4 | 0.6 | 1 |
| 1 | 0.5 | 0 |
| 6 | 0.3 | 1 |
| 3 | 0.2 | 0 |
| 5 | 0.2 | 1 |
| 2 | 0.1 | 0 |
| 7 | 0.0 | 0 |

Табл. 2

| id | > 0.25 | класс |
|----|--------|-------|
| 4 | 1 | 1 |
| 1 | 1 | 0 |
| 6 | 1 | 1 |
| 3 | 0 | 0 |
| 5 | 0 | 1 |
| 2 | 0 | 0 |
| 7 | 0 | 0 |

Табл. 3

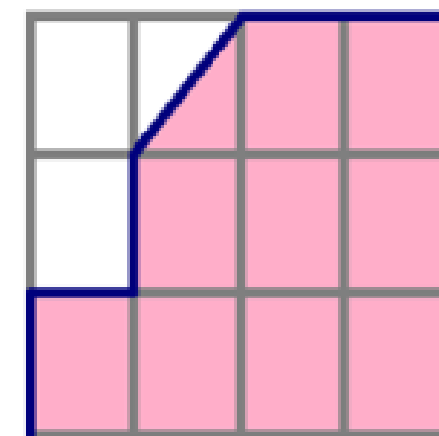
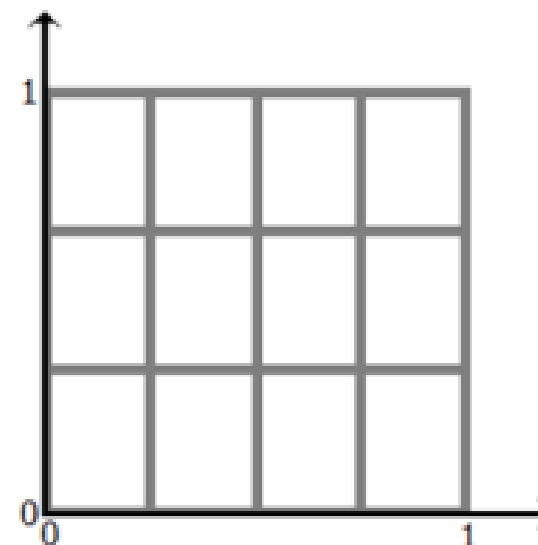
Метрики. ROC-AUC. Пример

Количество уникальных
TPR – количество «1».

Количество уникальных
FPR – количество «0».

Идем сверху вниз. Если
видим 1, то «вверх»,
если 0, то «вправо».
Если и 1, и 0, то идем
по диагонали.

| id | оценка | класс |
|----|--------|-------|
| 4 | 0.6 | 1 |
| 1 | 0.5 | 0 |
| 6 | 0.3 | 1 |
| 3 | 0.2 | 0 |
| 5 | 0.2 | 1 |
| 2 | 0.1 | 0 |
| 7 | 0.0 | 0 |



Метрики. ROC-AUC

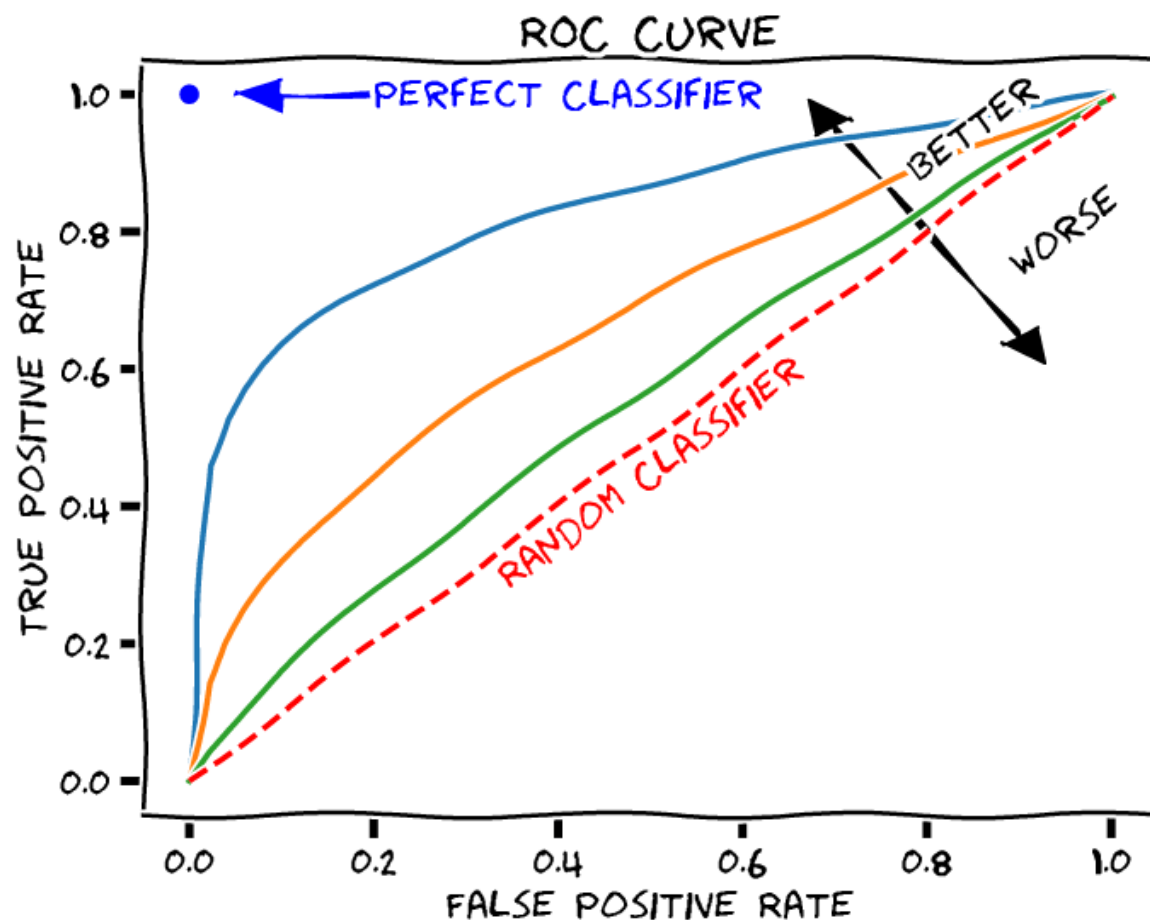
Значение ROC-AUC = 1 –
лучшая модель

ROC-AUC = 0.5 – случайная
модель

Чем выше значение, тем
выше ранжирующая
способность модели.

Интерпретация:

ROC-AUC показывает долю
пар (1, 0), которые
упорядочены корректно.

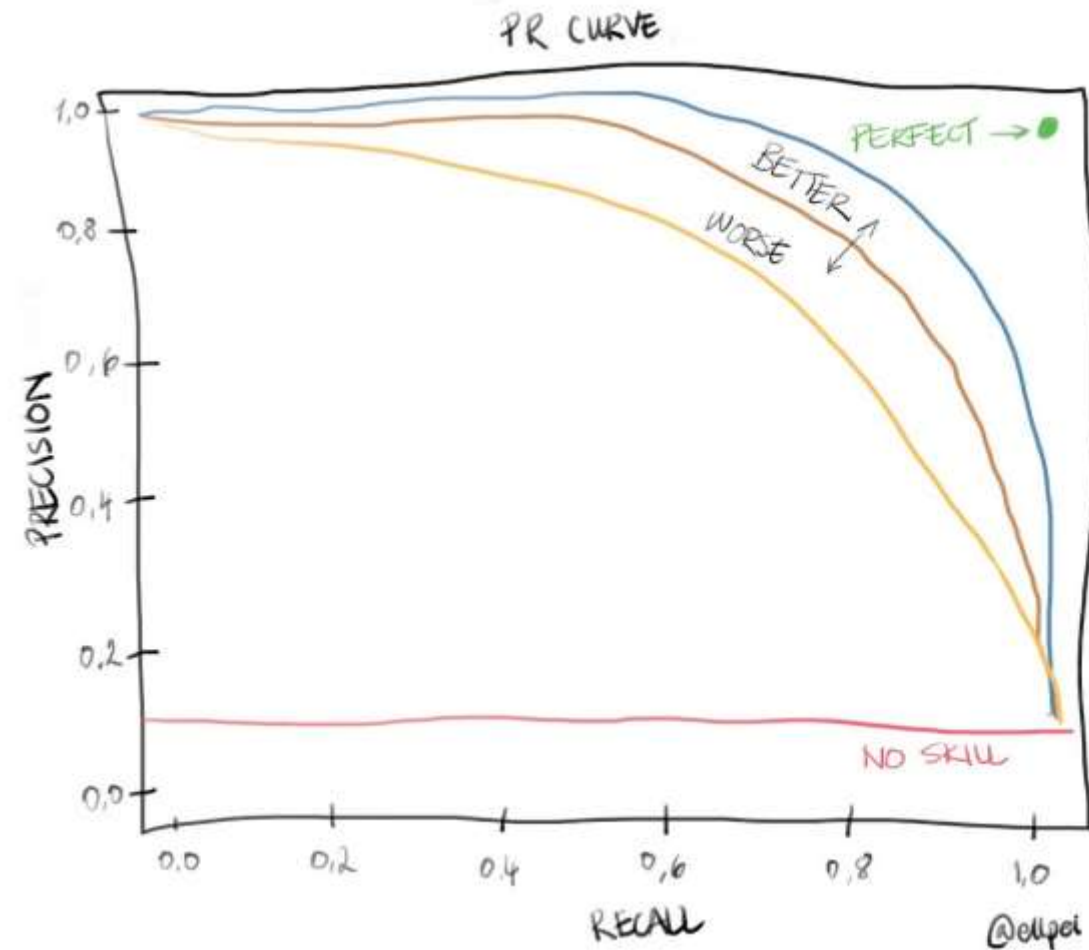


Метрики. Вновь Precision и Recall

Изменяя порог – меняем показатели Precision и Recall.

Чем выше порог, тем выше Precision и ниже Recall

Чем ниже порог, тем ниже Precision и выше Recall



Метрики. Регрессия

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

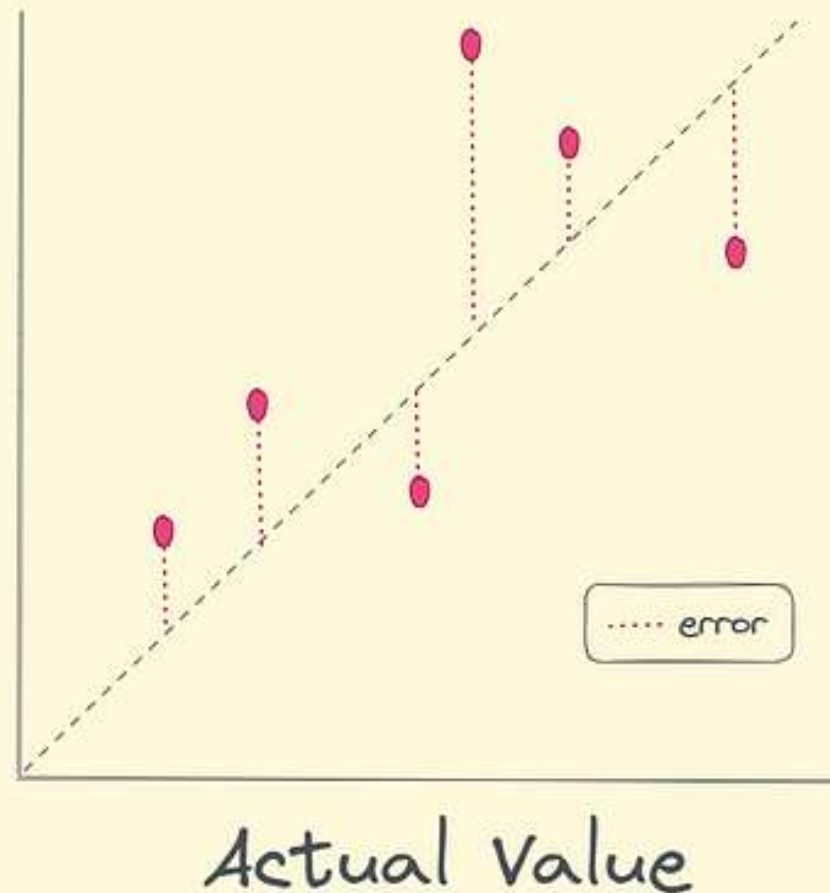
$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

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

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Predicted Value

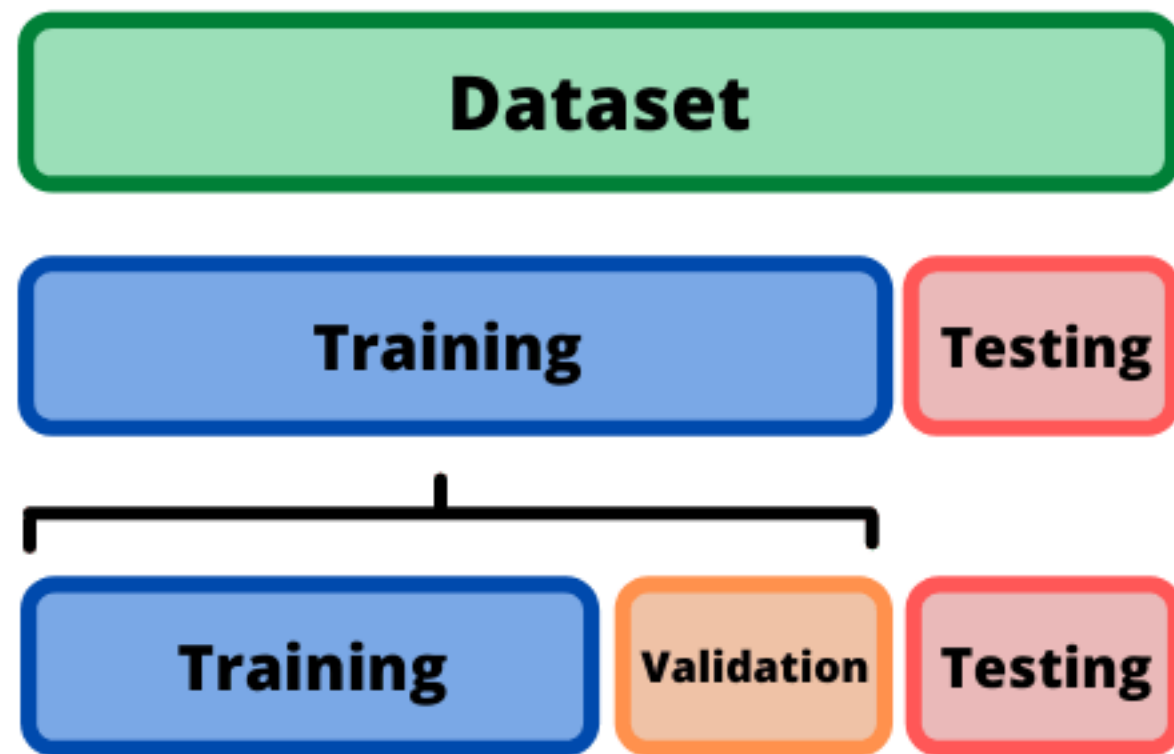


Методы валидации модели

Обучение на `train` части

Тестирование на `test` части

Подбор гиперпараметров на `validation` части

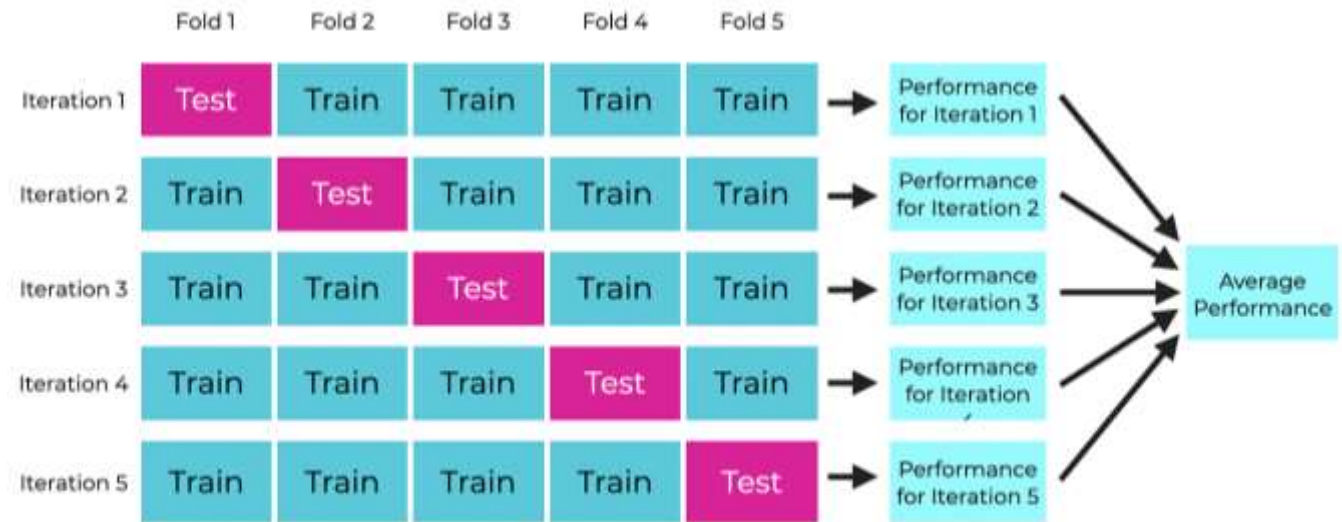


Кросс-валидация

Обучение на train части

Тестирование на test части

Подбор гиперпараметров на validation части **N раз** (например 5)



Time Series Cross Validation

