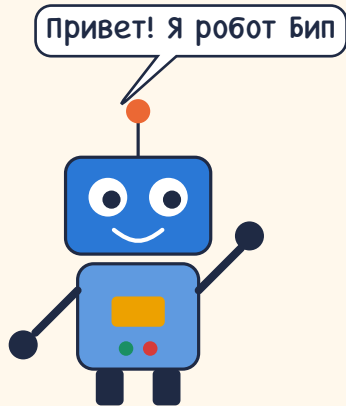


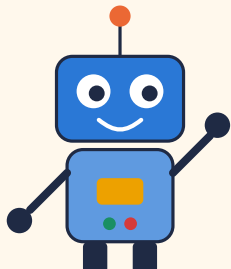
Как проверить, умный ли робот?

Оценка моделей машинного обучения
— простыми словами



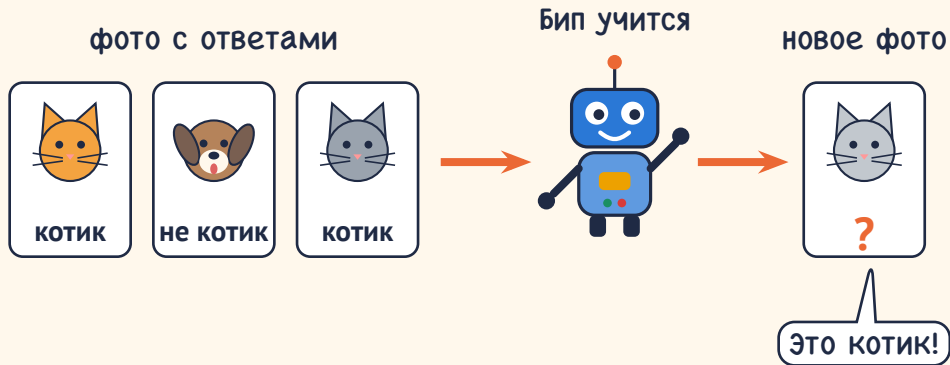
Знакомьтесь: робот Бип

Я учусь находить
котиков на фото!



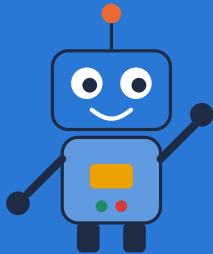
- Бип — это **программа**, которая учится на примерах.
- Мы показываем ему фото и говорим ответ: «**КОТИК**» или «**НЕ КОТИК**».
- Бип сам замечает приметы: уши-треугольники, усы, хвост.
- Потом Бип отвечает **сам** — на новых фото.

Как учится Бип



Учёные называют такого робота **моделью**, а фото с ответами — **данными**.

Часть 1
Честный экзамен для
робота



Можно ли проверять на тех же фото?

✗ Контрольная = задачи из учебника

Ученик мог просто **запомнить ответы**. Пятёрка ничего не говорит о том, понял ли он тему.

✓ Контрольная с новыми задачами

Запомнить ответы невозможно. Видно, **понял** ли ученик тему.

Робота проверяют на фото,
которые он ещё не видел

Делим 100 фото на три кучки



не открывать!

Учебник — 60 фото

Бип учится по этим фото

Тренировка
20 фото

пробуем,
исправляем,
выбираем
лучшего робота

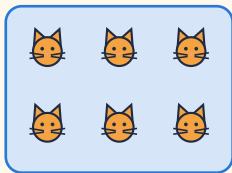
Контрольная
20 фото

открываем
один раз, в
самом конце

Как у повара. Повар пробует суп, пока готовит, и досаливает — это **тренировка**. Гости пробуют суп один раз, когда он готов, — это **контрольная**.

Перемешиваем честно

✗ Нечестно

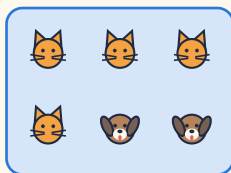


учебник



контрольная

✓ Честно



учебник

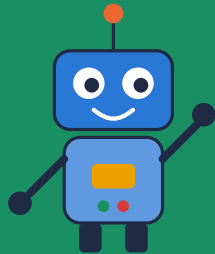


контрольная

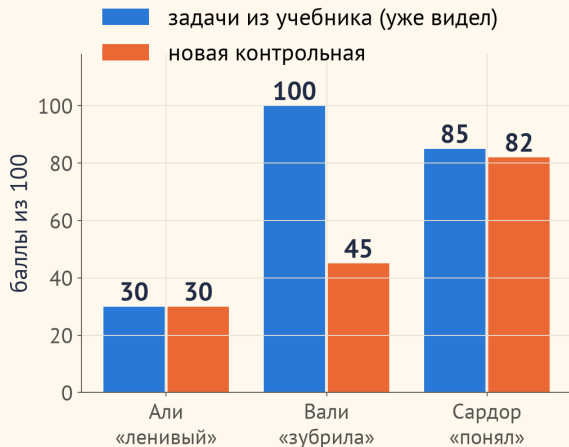
Бип ни разу не видел собак — на контрольной он растеряется.

В каждой кучке есть и котики, и не котики — примерно поровну, как во всём альбоме.

часть 2
Ленивый, зубрила и
умница



Три ученика готовятся к контрольной



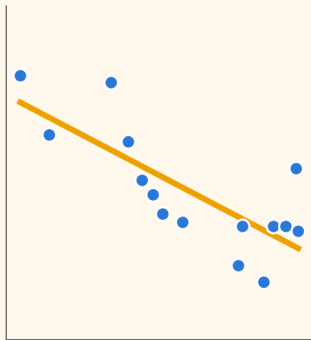
Али выучил одно правило: «не знаешь — пиши Б». Плохо и в учебнике, и на контрольной.

Вали выучил наизусть все ответы. В учебнике — 100, а новые задачи решить не может.

Сардор разобрался в теме. Хорошо везде!

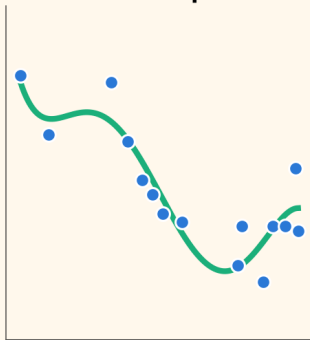
С роботами — то же самое

Ленивая линия



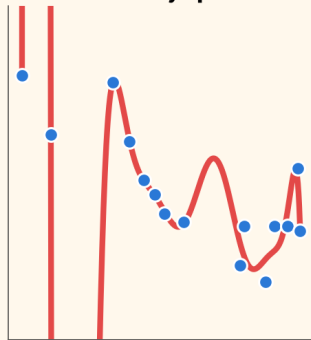
Недообучение

В самый раз!



В самый раз

Линия-зубрила

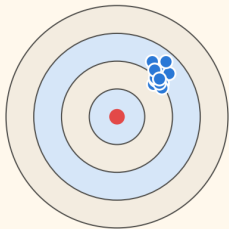


Переобучение

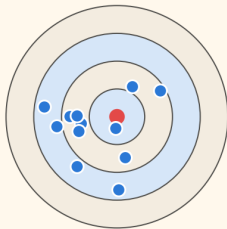
Синие точки — примеры из учебника. Робот рисует линию, которая должна подходить и к **новым** точкам.

Игра в дартс

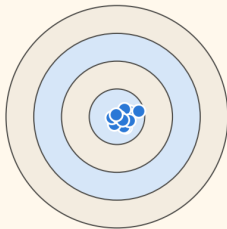
Кучно, но мимо
(ленивый)



Вокруг центра, но вразброс
(зубрила)



Метко и кучно
(понял)



Каждая синяя точка — робот, которого учили на **других** фото. Центр — правильный ответ.

Ленивый всегда промахивается в одну сторону. Зубрила каждый раз попадает в новое место.

Как помочь роботу?

Ленивый робот

плохо и на учебных, и на новых фото

- сделать робота «умнее» и сложнее
- подсказать больше примет
- **больше фото не поможет**

Робот-зубрила

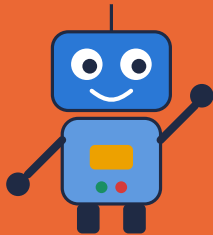
отлично на учебных, плохо на новых

- показать **больше разных фото**
- сделать робота проще
- всегда проверять на новых фото

Подсказка: сравни оценку в учебнике и на тренировке. Большая разница — зубрила!

Часть 3

Считаем ошибки бипа

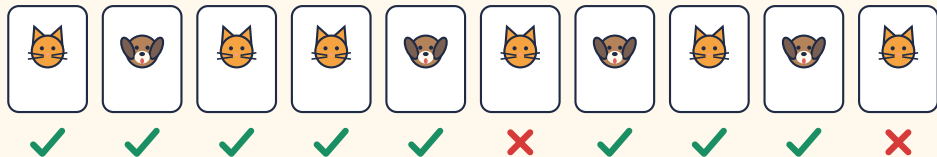


Четыре коробки ответов

	Бип сказал: «КОТИК»	Бип сказал: «НЕ КОТИК»
На фото КОТИК	 ✓ Нашёл котика! TP — True Positive	 ✗ Пропустил котика FN — False Negative
На фото не котик	 ✗ Ложная тревога: собаку назвал КОТИКОМ FP — False Positive	 ✓ Верно: это не КОТИК TN — True Negative

Зелёные коробки — правильные ответы, оранжевые — ошибки. Эта таблица называется **матрицей ошибок**.

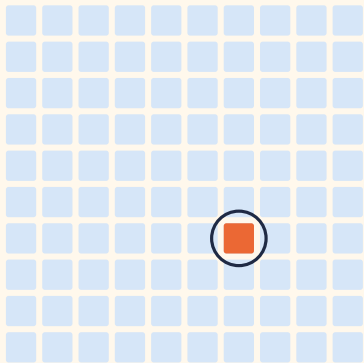
Accuracy: сколько ответов правильные?



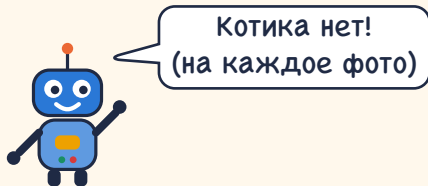
Бип ответил на 10 фото. Правильно — 8 раз, ошибся — 2 раза.

$$\text{Accuracy} = \frac{\text{правильные ответы}}{\text{все ответы}} = \frac{8}{10} = 80\%$$

Ловушка! Робот-лентяй получает 99 %



100 фото: только **1 котик**

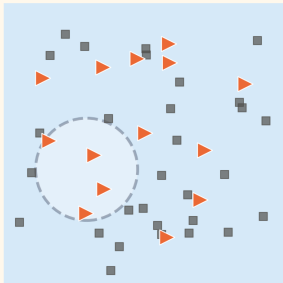


Правильно: 99 из 100 $\Rightarrow$ **99 %!**
Но Бип **не нашёл ни одного котика.**

Запомни: если котиков очень мало, ассигасу обманывает. Нужны другие оценки!

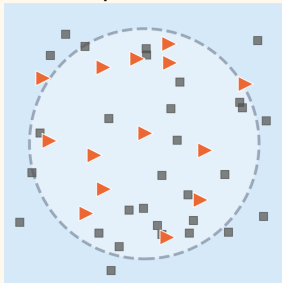
Рыбалка: Precision и Recall

Маленькая сеть



Поймали рыбок: 4 из 14
Мусора в сети: 0

Огромная сеть



Поймали рыбок: 12 из 14
Мусора в сети: 21

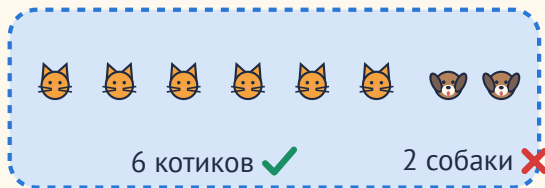
Рыбки — это то, что мы ищем, а серые квадраты — мусор. Сеть — это всё, что робот назвал «рыбкой».

Precision (точность):
сколько в сети **рыбы**, а не
мусора?

Recall (полнота):
сколько рыбок **из всего**
пруда поймали?

Считаем на котиках

Бип сказал «котик» — 8 фото



пропустил



Всего в альбоме 10 котиков.

$$\text{Precision} = \frac{6}{8} = 75\%$$

Когда Бип говорит «котик», он прав в 3 случаях из 4.

$$\text{Recall} = \frac{6}{10} = 60\%$$

Бип нашёл 6 котиков из 10.

Что важнее?



Пожарная сигнализация

Лучше лишний раз пискнуть от пара чайника, чем пропустить пожар.

Важнее Recall



Фильтр спама

Нельзя спрятать в «Спам» письмо от бабушки. Лучше пропустить одну рекламу.

Важнее Precision

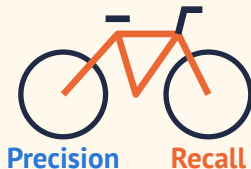
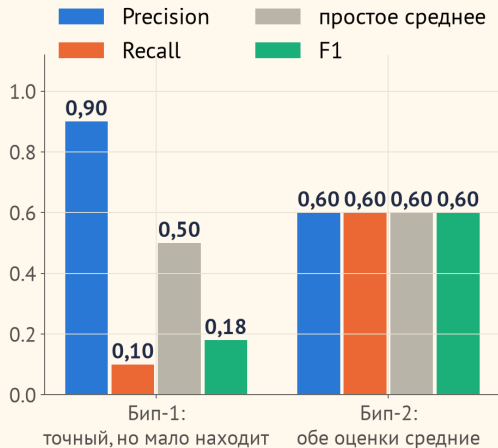


Врач и анализ

Если анализ ошибся — проверят ещё раз. А пропустить болезнь опасно.

Важнее Recall

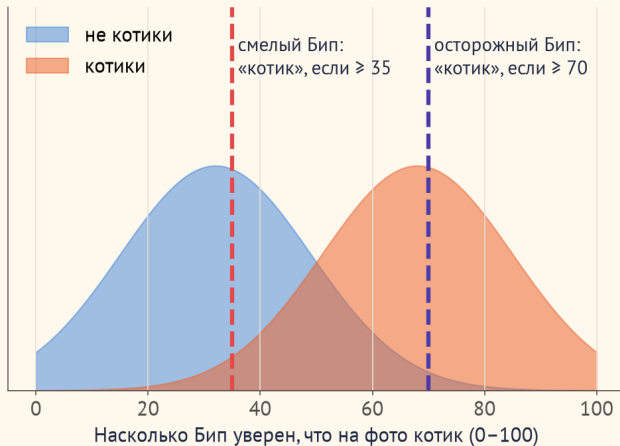
F1: велосипед с двумя колёсами



F1 — одна оценка из двух. Она высокая, только если **обе** оценки хорошие.

Бип-1 очень точный (0,9), но почти никого не находит (0,1). Простое среднее — 0,5, а F1 честно говорит: **0,18**.

Насколько Бип должен быть уверен?

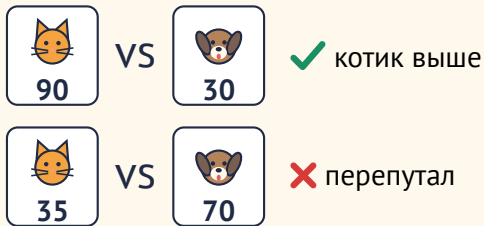


Смелый Бип (порог 35):
находит 97 котиков из
100, но 43 «не котика»
из 100 тоже называет
котиками.

Осторожный Бип (порог
70): почти не ошибается (1
из 100), но находит только
45 котиков из 100.

Порог выбираем **мы** — смотря что важнее: найти всех или не ошибаться.

ROC-AUC: игра в пары



Числа — баллы «это котик» от Бипа (0–100).

Берём **котика** и **не котика**. Кому Бип дал больше баллов?

AUC — доля пар, в которых котик получил больше баллов.

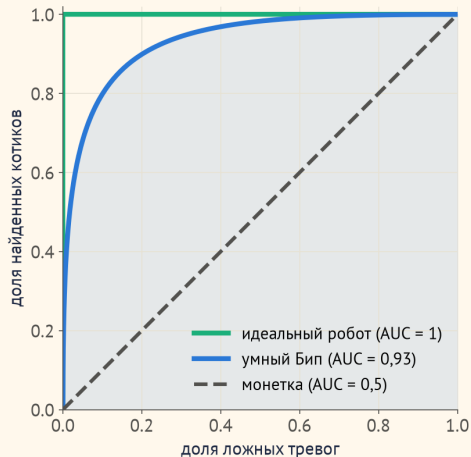
Пример: 20 правильных пар из 24

$$\Rightarrow AUC = \frac{20}{24} \approx 0,83$$

0,5 — Бип угадывает, как монетка.

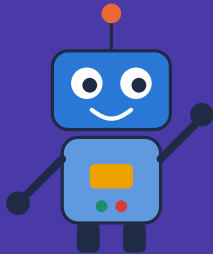
1 — котик всегда выше. Идеально!

ROC-кривая



- Каждая точка кривой — Бип с **одним порогом** уверенности.
- Вверх — нашли больше котиков. Вправо — больше ложных тревог.
- Чем ближе кривая к **левому верхнему углу**, тем лучше робот.
- **AUC** — площадь под кривой. Та же доля «правильных пар».

Часть 4
Повторим и проверим
себя



Шпаргалка



Слово	Детский вопрос
Accuracy	Сколько всех ответов правильные?
Precision	Когда робот говорит «котик», можно ли ему верить?
Recall	Сколько котиков робот нашёл?
F1	Хороши ли обе оценки сразу?
ROC-AUC	Ставит ли робот котиков выше не-котиков?
Недообучение	Ленивый: плохо везде
Переобучение	Зубрила: отлично на знакомом, плохо на новом

Проверь себя!



1. Робот на учебных фото набрал 100 %, а на новых — 50 %. Он **ленивый** или **зубрила**?
2. В альбоме 1000 фото и только 5 котиков. Робот всегда говорит «не котик». Какая у него accuracy? Хороший ли он?
3. Бип назвал котиками 10 фото, из них 5 — правда котики. Всего котиков 20. Найди Precision и Recall.
4. Для пожарной сигнализации что важнее: Precision или Recall?

Ответы



1. **Зубрила** (переобучение): отлично на знакомом, плохо на новом.
2. Accuracy = $\frac{995}{1000} = 99,5 \%$. Но робот **плохой**: он не нашёл ни одного котика.
3. Precision = $\frac{5}{10} = 50 \%$, Recall = $\frac{5}{20} = 25 \%$.
4. **Recall**: пропустить пожар намного хуже, чем лишний раз пискнуть.

Спасибо!

Вопросы?

