

Современные методы анализа данных и машинного обучения

Тема 6. Лекция 8

Классическое машинное обучение. Обучение с учителем.
Классификация

Юрий Саночкин

ysanochkin@hse.ru

НИУ ВШЭ, 2024

Машинное обучение

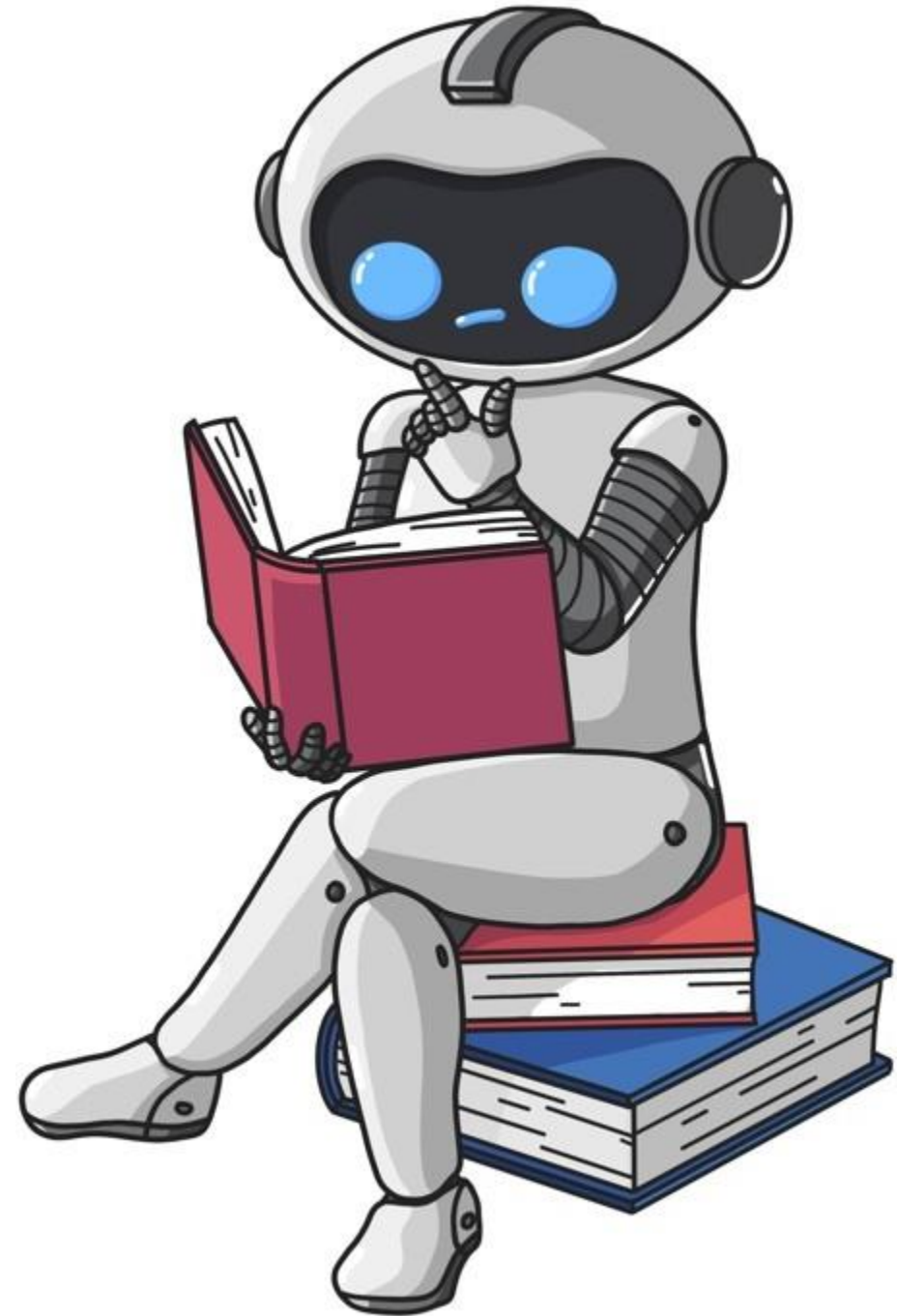
- Конечно, мы сейчас будем повторять это уже минимум в третий раз... но это такая база, что мне о-о-очень важно, чтобы вы точно это запомнили!

Машинное обучение

- Конечно, мы сейчас будем повторять это уже минимум в третий раз... но это такая база, что мне о-о-очень важно, чтобы вы точно это запомнили!
- Итак, вспомним еще раз (снова), в чём заключается концепция классического машинного обучения!

Машинное обучение

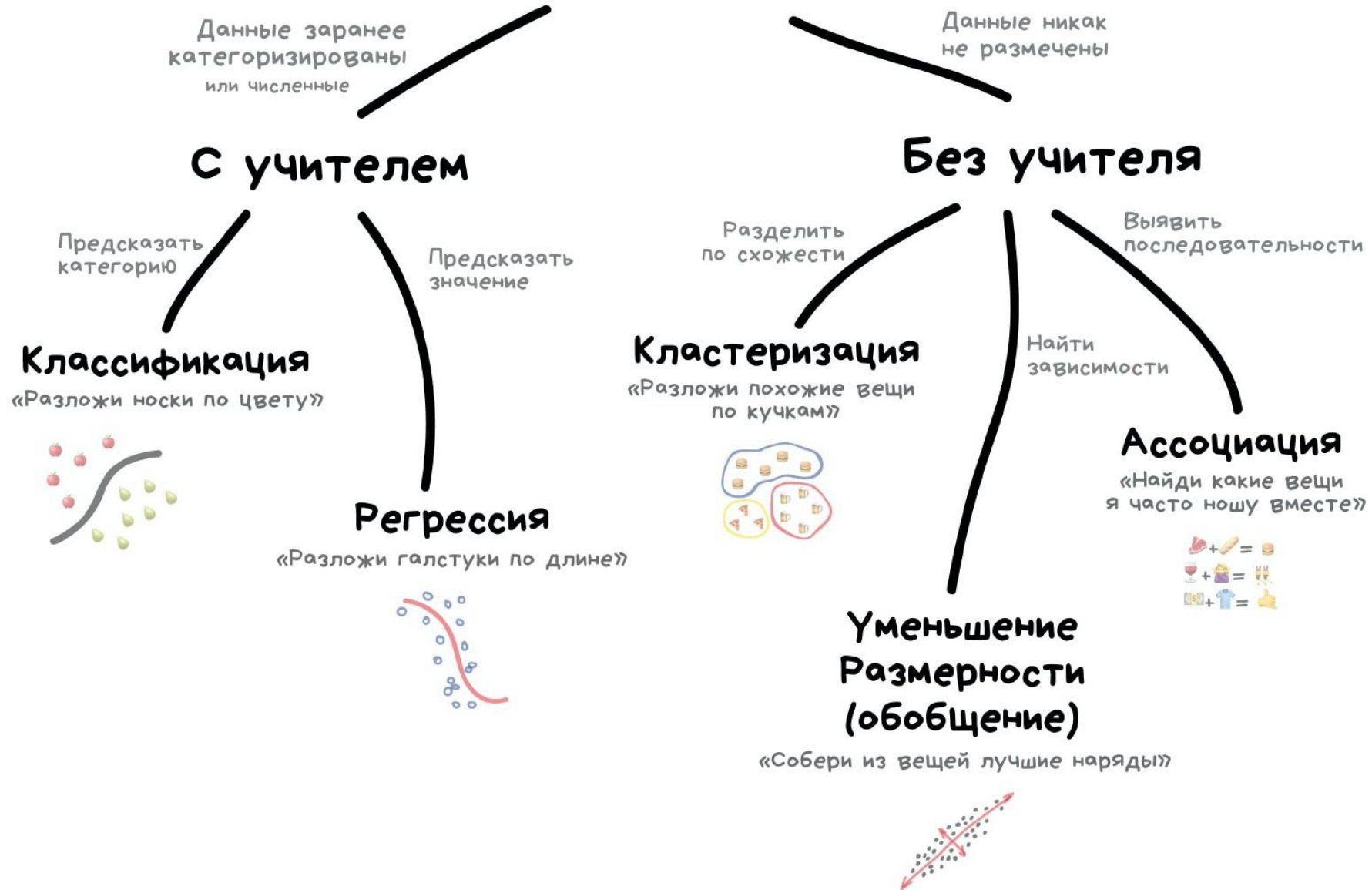
Наука о поиске
закономерностей в данных
с помощью компьютера и
математики



Машинное обучение

- На какие два больших блока можно разделить задачи классического машинного обучения?

Классическое Обучение



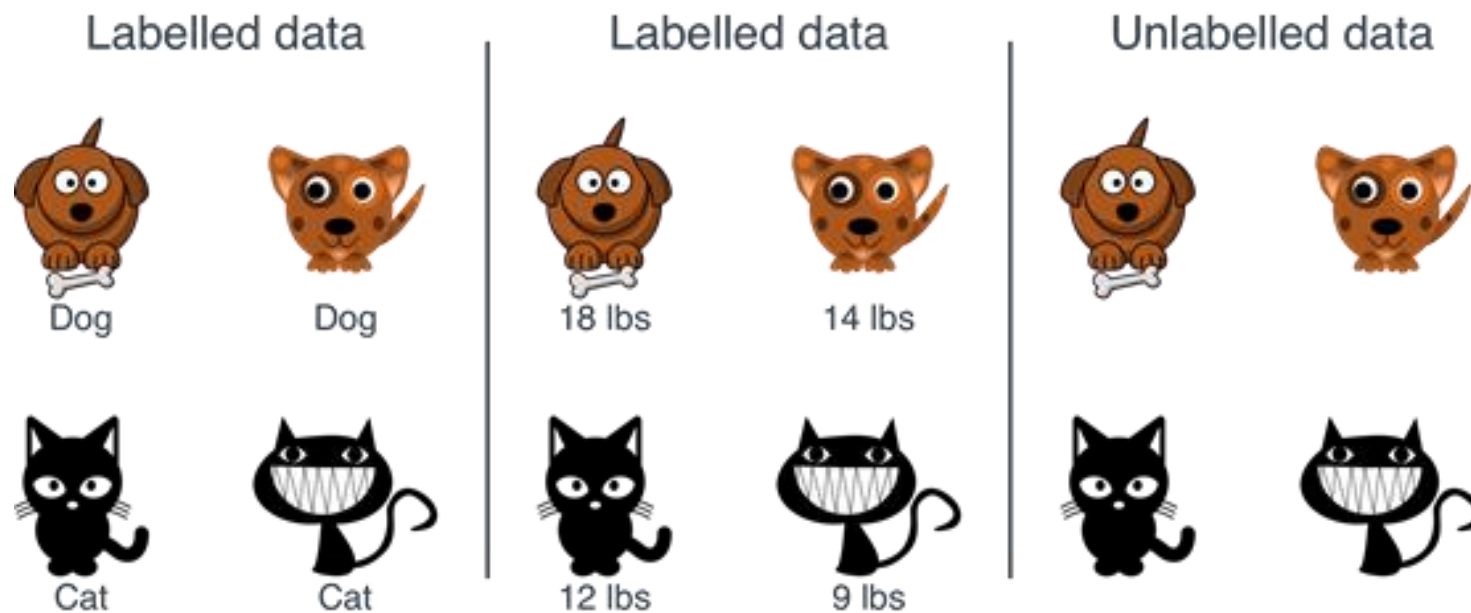
Классическое Обучение



Машинное обучение

- Данные блоки задач машинного обучения неразрывно связаны с понятием размеченных/неразмеченных данных

Размеченные (labelled) vs неразмеченные (unlabelled) данные



Классическое Обучение



Классическое Обучение



О регрессии мы уже говорили подробно, ну а сегодня – речь пойдет о классификации!

Обучение с учителем

- Еще раз повторим основные обозначения!

Обучение с учителем

- Еще раз повторим основные обозначения!
- X — множество всех объектов в пространстве признаков
- Y — область значений целевой переменной

Обучение с учителем

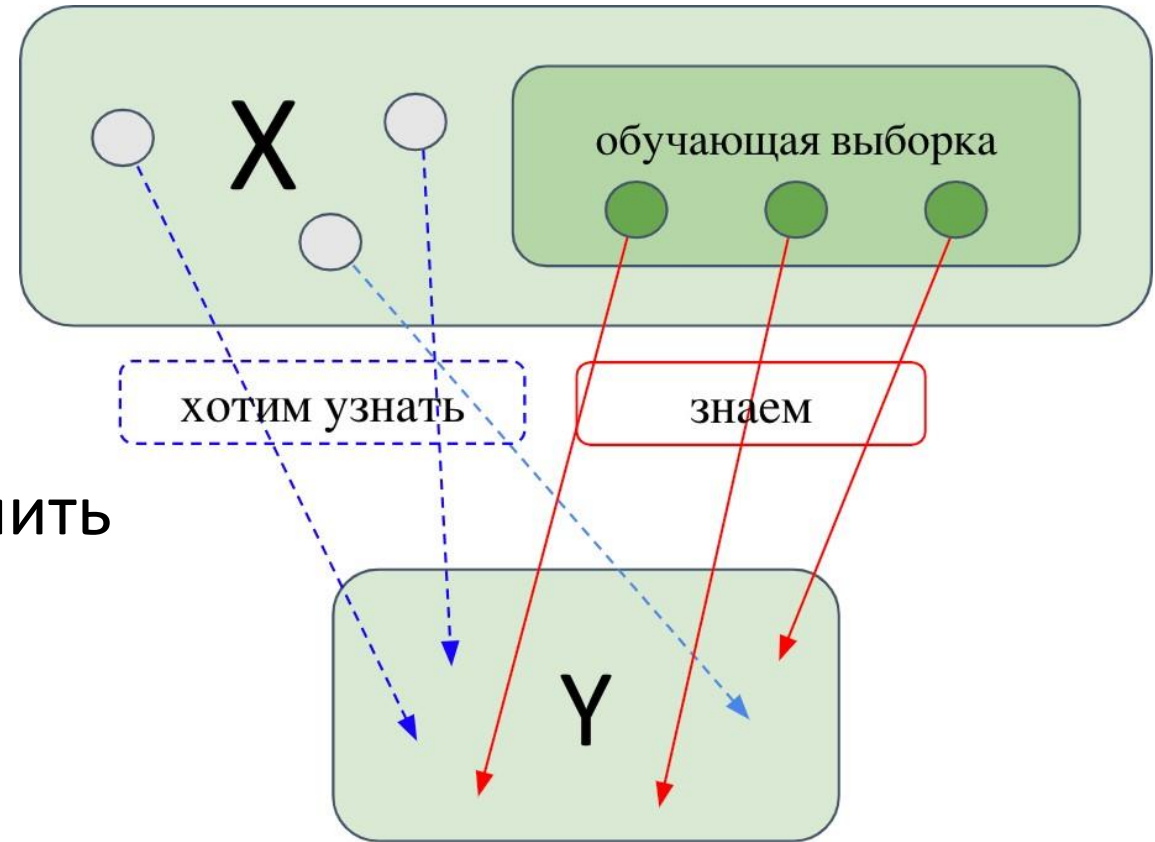
- Еще раз повторим основные обозначения!
- X — множество всех объектов в пространстве признаков
- Y — область значений целевой переменной
- Что представляет собой машинное обучение в этих терминах?
- Фактически это про поиск неизвестной зависимости:
- $f: X \rightarrow Y$ — неизвестная закономерность, функция
- Может даже иметь стохастическую природу!

Обучение с учителем

- Как мы это осуществляем?
- Дано: Обучающая выборка вида $\{(X_i, y_i)\}_{i=1}^n$
- Цель: Максимально точно определить и приблизить f .

Обучение с учителем

- Как мы это осуществляем?
- Дано: Обучающая выборка вида $\{(X_i, y_i)\}_{i=1}^n$
- Цель: Максимально точно определить и приблизить f .



Задача классификации

- Что такое задача классификации?
- Говоря по простому — задача, где мы хотим предсказать принадлежность объекта к одному из заранее определенных классов (категорий)

Задача классификации

- Что такое задача классификации?
- Говоря по простому — задача, где мы хотим предсказать принадлежность объекта к одному из заранее определенных классов (категорий)
- Приведите примеры каких-нибудь задач классификации
 - Предсказание оттока клиентов/сотрудников на основе их поведения
 - Классификация клеток ткани на здоровые и опухолевые
 - Детекция объектов на фото
 - И так далее

Метрики качества и функционал ошибки

Метрики качества и функционал ошибки

- В прошлый раз мы уже обсуждали с вами понятия метрик качества и функционала ошибки.
- Напомните, в чем разница между ними?

Метрики качества и функционал ошибки

- Метрики качества используются для непосредственной оценки качества обученного алгоритма, с учетом наших бизнес-потребностей
- Проще говоря: смотрим на полученный моделью результат и сравниваем его с правильными ответами
- Функционал ошибки же оценивает некую математическую функцию, которую в процессе обучения пытается минимизировать модель

Метрики качества и функционал ошибки

- И, как мы и говорили в прошлый раз, к сожалению, с метриками качества для классификации — всё обстоит не так радужно, как с регрессией.
- В каком смысле?

Метрики качества и функционал ошибки

- И, как мы и говорили в прошлый раз, к сожалению, с метриками качества для классификации — всё обстоит не так радужно, как с регрессией.
- В каком смысле?
- В том смысле, что:
 - Во-первых, метрик классификации очень много — гораздо больше, чем для регрессии — и в разных случаях будут использоваться разные из них. Нужно понимать преимущества и недостатки каждой.

Метрики качества и функционал ошибки

- И, как мы и говорили в прошлый раз, к сожалению, с метриками качества для классификации — всё обстоит не так радужно, как с регрессией.
- В каком смысле?
- В том смысле, что:
 - Во-вторых, метрики классификации очень часто не совпадают с функционалом ошибки, поскольку попросту не могут быть минимизированы/максимизированы методами математической оптимизации напрямую. Как следствие часто требуют аппроксимацию другими математическими функциями.

Метрики качества и функционал ошибки

- И, как мы и говорили в прошлый раз, к сожалению, с метриками качества для классификации — всё обстоит не так радужно, как с регрессией.
- В каком смысле?
- Всё это, увы, непросто, но мы попробуем разобраться!

Матрица ошибок

- Для работы со многими метриками качества классификации нам потребуется такое важное понятие как «матрица ошибок» (confusion matrix)

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP	FN
	NEGATIVE (0)	FP	TN

Матрица ошибок

- Для работы со многими метриками качества классификации нам потребуется такое важное понятие как «матрица ошибок» (confusion matrix)
- Обратите, пожалуйста, внимание на обозначения в таблице. Они — общепринятые, и их нужно очень хорошо знать!

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP	FN
	NEGATIVE (0)	FP	TN

Матрица ошибок

- Итак, какие же метрики качества по матрице ошибок для классификации вы знаете?

Матрица ошибок

- Итак, какие же метрики качества по матрице ошибок для классификации вы знаете?
- Для полноты картины начнем с Accuracy или доли верно угаданных ответов.

Матрица ошибок

- Итак, какие же метрики качества по матрице ошибок для классификации вы знаете?
- Для полноты картины начнем с Accuracy или доли верно угаданных ответов.
- Формульно она выглядит следующим образом:

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

Матрица ошибок

- Итак, какие же метрики качества по матрице ошибок для классификации вы знаете?
- Для полноты картины начнем с Accuracy или доли верно угаданных ответов.
- Формульно она выглядит следующим образом:

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

- Посчитаем руками для следующей матрицы ошибок:

—	True Class 1	True Class 0
Predict Class 1	7	0
Predict Class 0	8	435

Матрица ошибок

- Итак, какие же метрики качества по матрице ошибок для классификации вы знаете?
- Для полноты картины начнем с Accuracy или доли верно угаданных ответов.
- Формульно она выглядит следующим образом:

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

Какие проблемы могут быть у Accuracy? Почему её одной недостаточно?

- Посчитаем руками для следующей матрицы ошибок:

—	True Class 1	True Class 0
Predict Class 1	7	0
Predict Class 0	8	435

Матрица ошибок

- Следующая метрика – Precision или точность.

Матрица ошибок

- Следующая метрика – Precision или точность.
- Precision – это доля верно угаданных элементов положительного класса среди всех элементов, помеченных положительными.
- Данная метрика демонстрирует способность модели отличать класс от других классов.

Матрица ошибок

- Следующая метрика – Precision или точность.
- Precision – это доля верно угаданных элементов положительного класса среди всех элементов, помеченных положительными.
- Данная метрика демонстрирует способность модели отличать класс от других классов.
- Формульно она выглядит следующим образом:

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

Матрица ошибок

- Формульно она выглядит следующим образом:

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

- Когда эта метрика максимальна/минимальна? В каких диапазонах могут лежать ее значения?

Матрица ошибок

- Формульно она выглядит следующим образом:

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

- Когда эта метрика максимальна/минимальна? В каких диапазонах могут лежать ее значения?
- Посчитаем руками для следующей матрицы ошибок:

—	True Class 1	True Class 0
Predict Class 1	7	0
Predict Class 0	8	435

Матрица ошибок

- Следующая – Recall или полнота.

Матрица ошибок

- Следующая – Recall или полнота.
- Recall – это доля верно помеченных объектов положительного класса среди всех объектов положительного класса.
- Данная метрика демонстрирует способность алгоритма обнаруживать данный класс вообще.

Матрица ошибок

- Следующая – Recall или полнота.
- Recall – это доля верно помеченных объектов положительного класса среди всех объектов положительного класса.
- Данная метрика демонстрирует способность алгоритма обнаруживать данный класс вообще.
- Формульно она выглядит следующим образом:

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

Матрица ошибок

- Формульно она выглядит следующим образом:

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

- Когда эта метрика максимальна/минимальна? В каких диапазонах могут лежать ее значения?

Матрица ошибок

- Формульно она выглядит следующим образом:

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

- Когда эта метрика максимальна/минимальна? В каких диапазонах могут лежать ее значения?
- Посчитаем руками для следующей матрицы ошибок:

—	True Class 1	True Class 0
Predict Class 1	7	0
Predict Class 0	8	435

Матрица ошибок

- Впрочем, у предыдущих метрик есть важный недостаток: каждая из них демонстрирует качество работы очень разных аспектов нашего классификатора.

Матрица ошибок

- Впрочем, у предыдущих метрик есть важный недостаток: каждая из них демонстрирует качество работы очень разных аспектов нашего классификатора.
- Но как нам контролировать качество алгоритма не двумя или тремя метриками, а всего одной, достаточно всеобъемлющей? Это вообще возможно?

Матрица ошибок

- Впрочем, у предыдущих метрик есть важный недостаток: каждая из них демонстрирует качество работы очень разных аспектов нашего классификатора.
- Но как нам контролировать качество алгоритма не двумя или тремя метриками, а всего одной, достаточно всеобъемлющей? Это вообще возможно?
- Ну, мы по крайней мере можем постараться создать такую метрику!

Матрица ошибок

- F1-score или f1-мера.

Матрица ошибок

- F1-score или f1-мера.
- F1-score – это среднее гармоническое точности и полноты.
- Данная метрика одинаково и сбалансированно учитывает важность точности и полноты, а потому показывает более четкую картину работы нашего алгоритма.

Матрица ошибок

- F1-score или f1-мера.
- F1-score – это среднее гармоническое точности и полноты.
- Данная метрика одинаково и сбалансированно учитывает важность точности и полноты, а потому показывает более четкую картину работы нашего алгоритма.
- Формульно она выглядит следующим образом:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Матрица ошибок

- Формульно она выглядит следующим образом:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

- Когда эта метрика максимальна/минимальна? В каких диапазонах могут лежать ее значения?

Матрица ошибок

- Формульно она выглядит следующим образом:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

- Когда эта метрика максимальна/минимальна? В каких диапазонах могут лежать ее значения?
- Посчитаем руками для следующей матрицы ошибок:

—	True Class 1	True Class 0
Predict Class 1	7	0
Predict Class 0	8	435

Другие метрики качества

- ROC-AUC или площадь под ROC кривой.

Другие метрики качества

- ROC-AUC или площадь под ROC кривой.
- ROC-AUC – непростая и отличающаяся от других мера. Данная мера показывает, насколько точно упорядочены элементы выборки по своим вероятностям (скорам).

Алгоритм построения ROC-кривой

Алгоритм построения ROC-кривой

1. Делаем предсказания моделью и получаем вероятности принадлежности объекта к положительному классу.

Алгоритм построения ROC-кривой

1. Делаем предсказания моделью и получаем вероятности принадлежности объекта к положительному классу.

Стоп, что? Откуда здесь вообще взялись какие-то вероятности?.. Мы же решаем задачу классификации, а значит класс предсказываем!..

Алгоритм построения ROC-кривой

1. Делаем предсказания моделью и получаем вероятности принадлежности объекта к положительному классу.
2. Делим отрезок $[0,1]$ на оси x на число частей, равное количеству элементов отрицательного класса; на оси y — на число частей, равное количеству элементов положительного класса.

Алгоритм построения ROC-кривой

1. Делаем предсказания моделью и получаем вероятности принадлежности объекта к положительному классу.
2. Делим отрезок $[0,1]$ на оси x на число частей, равное количеству элементов отрицательного класса; на оси y — на число частей, равное количеству элементов положительного класса.
3. Сортируем объекты по полученным вероятностям по убыванию.

Алгоритм построения ROC-кривой

1. Делаем предсказания моделью и получаем вероятности принадлежности объекта к положительному классу.
2. Делим отрезок $[0,1]$ на оси x на число частей, равное количеству элементов отрицательного класса; на оси y — на число частей, равное количеству элементов положительного класса.
3. Сортируем объекты по полученным вероятностям по убыванию.
4. Если есть элементы с одинаковыми вероятностями, то группируем их вместе.

Алгоритм построения ROC-кривой

1. Делаем предсказания моделью и получаем вероятности принадлежности объекта к положительному классу.
2. Делим отрезок $[0,1]$ на оси x на число частей, равное количеству элементов отрицательного класса; на оси y — на число частей, равное количеству элементов положительного класса.
3. Сортируем объекты по полученным вероятностям по убыванию.
4. Если есть элементы с одинаковыми вероятностями, то группируем их вместе.
5. Идем по таблице со сгруппированными вероятностями:

Алгоритм построения ROC-кривой

5. Идем по таблице со сгруппированными вероятностями:

Алгоритм построения ROC-кривой

5. Идем по таблице со сгруппированными вероятностями:
 - Если реальный класс рассматриваемого объекта положительный, то делаем шаг вверх;

Алгоритм построения ROC-кривой

5. Идем по таблице со сгруппированными вероятностями:
 - Если реальный класс рассматриваемого объекта положительный, то делаем шаг вверх;
 - Если реальный класс рассматриваемого объекта отрицательный, то делаем шаг вправо;

Алгоритм построения ROC-кривой

5. Идем по таблице со сгруппированными вероятностями:

- Если реальный класс рассматриваемого объекта положительный, то делаем шаг вверх;
- Если реальный класс рассматриваемого объекта отрицательный, то делаем шаг вправо;
- Если мы рассматриваем группу объектов, то делаем шаг по диагонали на нужное число клеток вправо и вверх.

Алгоритм построения ROC-кривой

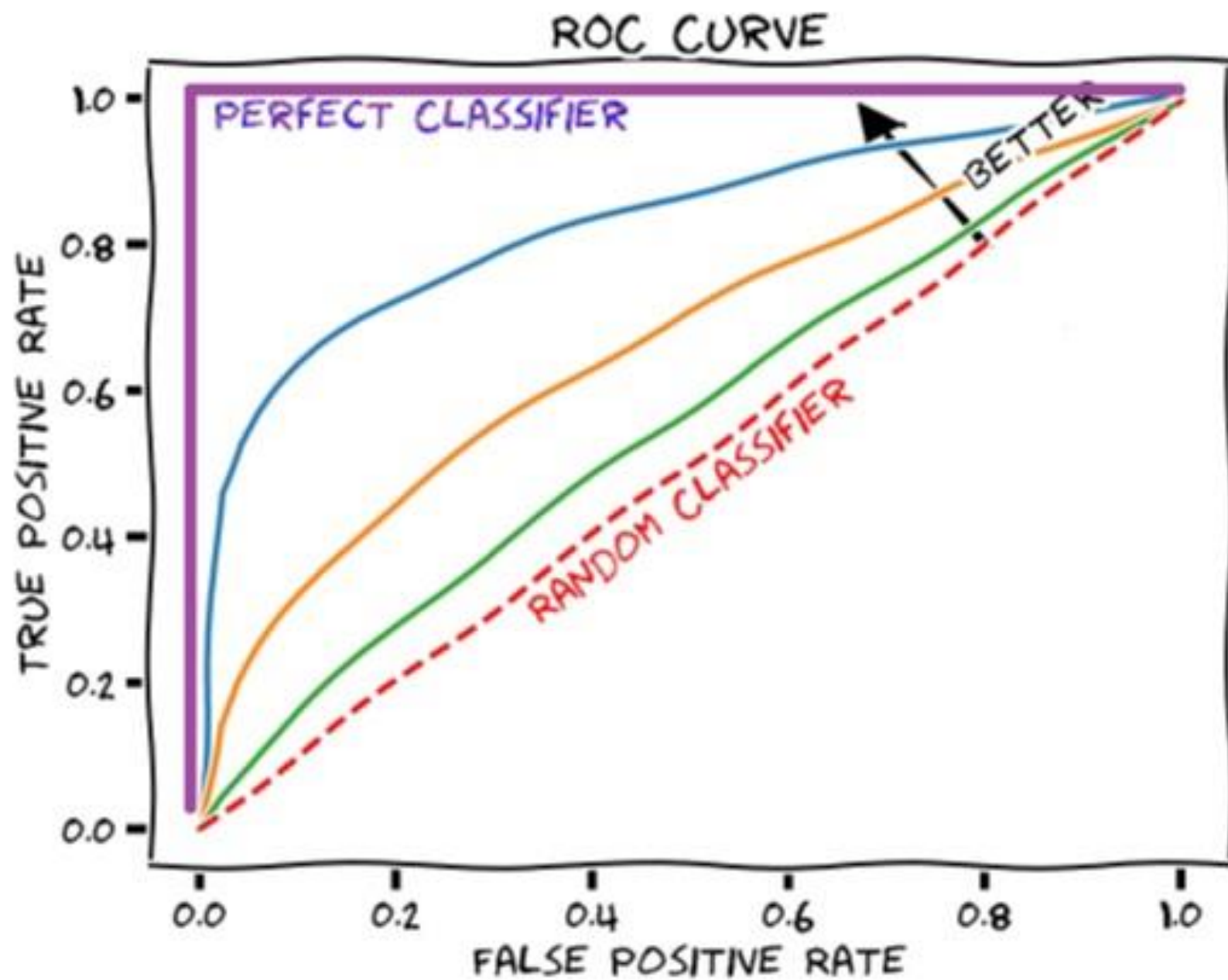
5. Идем по таблице со сгруппированными вероятностями:
 - Если реальный класс рассматриваемого объекта положительный, то делаем шаг вверх;
 - Если реальный класс рассматриваемого объекта отрицательный, то делаем шаг вправо;
 - Если мы рассматриваем группу объектов, то делаем шаг по диагонали на нужное число клеток вправо и вверх.
6. Получившаяся кривая выходит из $[0,0]$ и приходит в $[1,1]$. Площадь под ней и называется ROC-AUC.

Другие метрики качества

- ROC-AUC или площадь под ROC кривой.
- Данная мера показывает, насколько точно упорядочены элементы выборки по своим вероятностям (скорам).

Другие метрики качества

- ROC-AUC или площадь под ROC кривой.
- Данная мера показывает, насколько точно упорядочены элементы выборки по своим вероятностям (скорам).
- Когда эта метрика максимальна/минимальна? В каких диапазонах могут лежать ее значения?



Другие метрики качества

- ROC-AUC или площадь под ROC кривой.
- Данная мера показывает, насколько точно упорядочены элементы выборки по своим вероятностям (скорам).
- Когда эта метрика максимальна/минимальна? В каких диапазонах могут лежать ее значения?

Другие метрики качества

- ROC-AUC или площадь под ROC кривой.
- Данная мера показывает, насколько точно упорядочены элементы выборки по своим вероятностям (скорам).
- Когда эта метрика максимальна/минимальна? В каких диапазонах могут лежать ее значения?
- Давайте порассуждаем, почему метрика, построенная по такому алгоритму, достаточно устойчива к несбалансированным выборкам? Почему она показывает ранжирование объектов?

Другие метрики качества

- ROC-AUC или площадь под ROC кривой.
- Данная мера показывает, насколько точно упорядочены элементы выборки по своим вероятностям (скорам).
- Когда эта метрика максимальна/минимальна? В каких диапазонах могут лежать ее значения?
- Давайте порассуждаем, почему метрика, построенная по такому алгоритму, достаточно устойчива к несбалансированным выборкам? Почему она показывает ранжирование объектов?
- Мы потренируемся в расчете данной метрики на семинаре!

Линейные модели

Линейные модели

- Отвлечемся на время от разбора задач классификации (мы обязательно к ним скоро вернемся) и поговорим о моделях машинного обучения.
- Сколько моделей машинного обучения мы знаем на текущий момент? Что это за модели?

Линейные модели

- Отвлечемся на время от разбора задач классификации (мы обязательно к ним скоро вернемся) и поговорим о моделях машинного обучения.
- Сколько моделей машинного обучения мы знаем на текущий момент? Что это за модели?
- По сути, в рамках нашего курса мы разбирали всего лишь одну модель. Возможно, вы проходили на других курсах или слышали сами по себе и о других моделях; но на текущий момент — нам с вами тут знакома лишь модель KNN.

Линейные модели

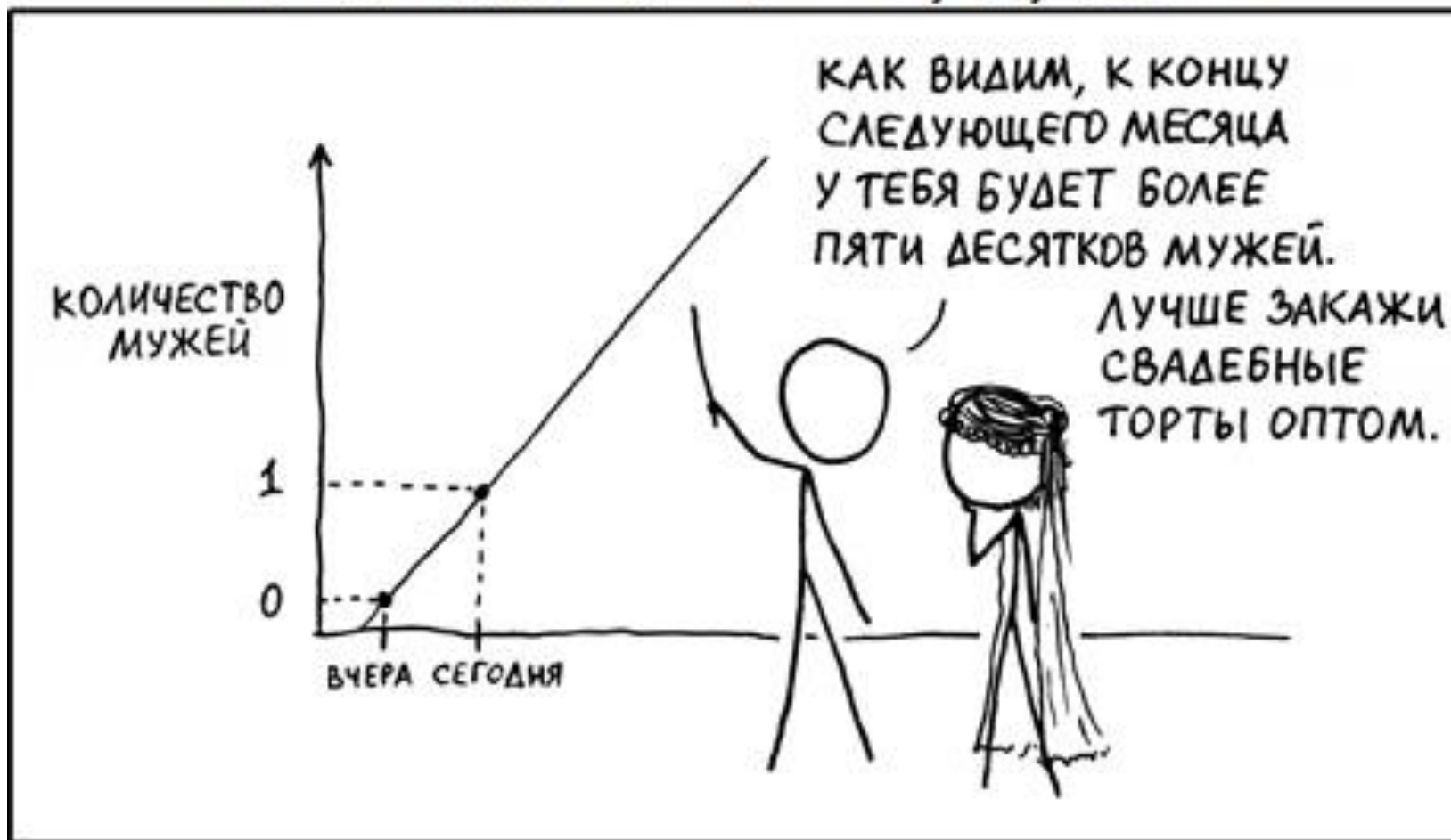
- Отвлечемся на время от разбора задач классификации (мы обязательно к ним скоро вернемся) и поговорим о моделях машинного обучения.
- Сколько моделей машинного обучения мы знаем на текущий момент? Что это за модели?
- По сути, в рамках нашего курса мы разбирали всего лишь одну модель. Возможно, вы проходили на других курсах или слышали сами по себе и о других моделях; но на текущий момент — нам с вами тут знакома лишь модель KNN.
- Кажется, самое время расширить наш арсенал! :)

Линейные модели

- Представляю вашему вниманию — линейные модели!

Линейные модели

Аналитик и линейная регрессия



Линейные модели

- Перейдем к постановке задачи линейных моделей. Вначале рассмотрим задачу регрессии (она и здесь проще).
- Пусть у нас есть данные, которые записаны привычным для нас образом, в виде матрицы признаков размера $M \times N$, где M – число объектов, а N – число признаков. Кроме того, у нас есть столбец таргета Y .

Линейные модели

- Перейдем к постановке задачи линейных моделей. Вначале рассмотрим задачу регрессии (она и здесь проще).
- Пусть у нас есть данные, которые записаны привычным для нас образом, в виде матрицы признаков размера $M \times N$, где M – число объектов, а N – число признаков. Кроме того, у нас есть столбец таргета Y .
- Мы хотим построить прямую (в общем случае гиперплоскость), которая бы наилучшим образом приближала наши данные в каждой из точек (для каждого объекта).

Линейные модели

- Перейдем к постановке задачи линейных моделей. Вначале рассмотрим задачу регрессии (она и здесь проще).
- Пусть у нас есть данные, которые записаны привычным для нас образом, в виде матрицы признаков размера $M \times N$, где M – число объектов, а N – число признаков. Кроме того, у нас есть столбец таргета Y .
- Мы хотим построить прямую (в общем случае гиперплоскость), которая бы наилучшим образом приближала наши данные в каждой из точек (для каждого объекта).

Вопрос на миллион: что такое гиперплоскость?

Линейные модели

- В целях построения «наилучшей» прямой, мы можем рассматривать в качестве приближения объекта прямой привычное нам выражение: $(\tilde{y}_i - y_i)^2$. Чем меньше это значение, тем лучше прямая приближает объект.

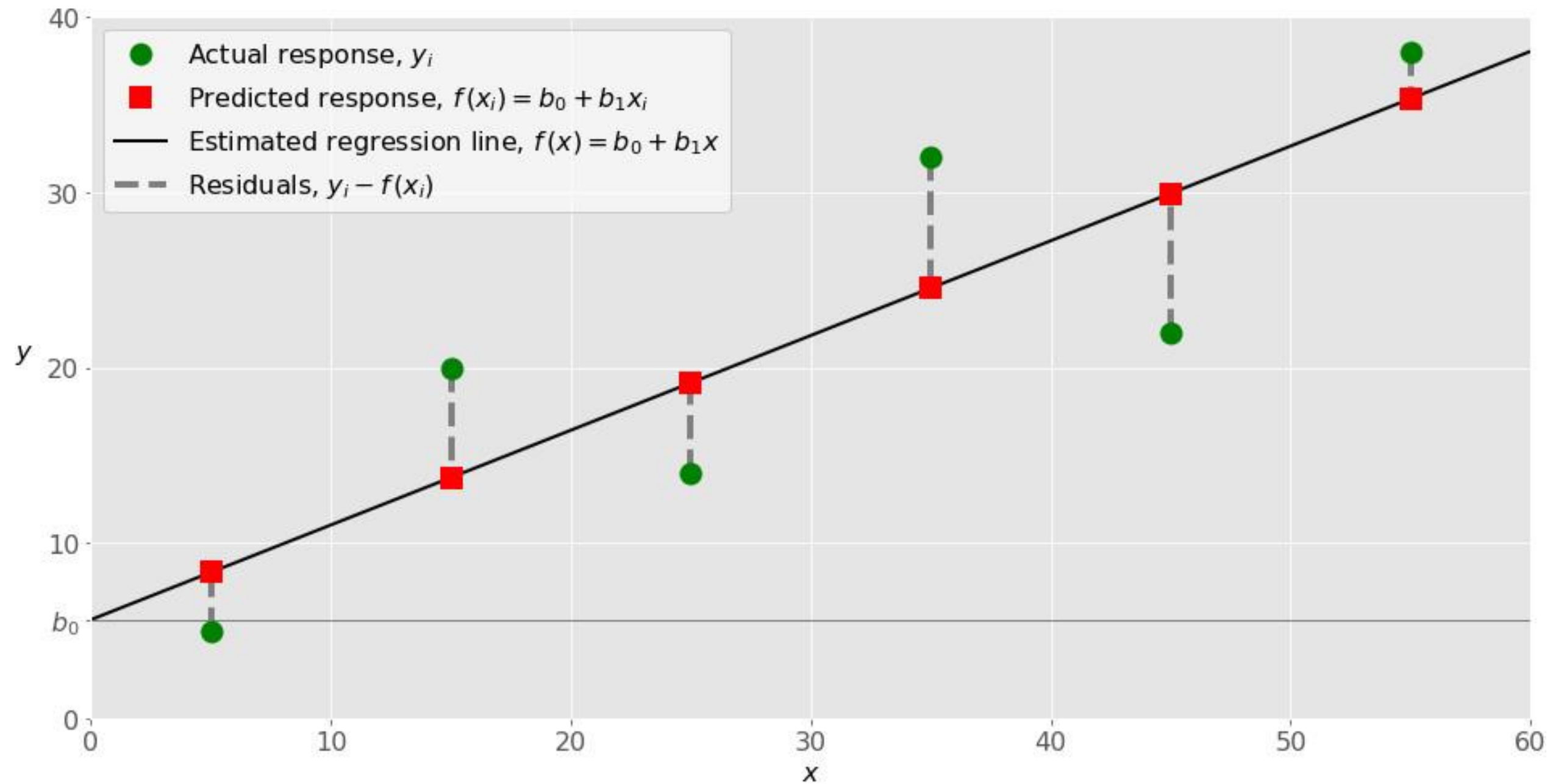
Линейные модели

- В целях построения «наилучшей» прямой, мы можем рассматривать в качестве приближения объекта прямой привычное нам выражение: $(\tilde{y}_i - y_i)^2$. Чем меньше это значение, тем лучше прямая приближает объект.
- Если же мы хотим приближать несколько объектов, то можно использовать обобщение: $\sum_i^n (\tilde{y}_i - y_i)^2$.

Линейные модели

- В целях построения «наилучшей» прямой, мы можем рассматривать в качестве приближения объекта прямой привычное нам выражение: $(\tilde{y}_i - y_i)^2$. Чем меньше это значение, тем лучше прямая приближает объект.
- Если же мы хотим приближать несколько объектов, то можно использовать обобщение: $\sum_i^n (\tilde{y}_i - y_i)^2$.

Кстати, что за обозначения тут используются?
Что есть что?



Линейные модели

- С тем, что мы бы хотели сделать с прямой, мы более-менее разобрались. Но давайте поговорим о самой прямой (или, опять же, гиперплоскости в общем случае).

Линейные модели

- С тем, что мы бы хотели сделать с прямой, мы более-менее разобрались. Но давайте поговорим о самой прямой (или, опять же, гиперплоскости в общем случае).
- Из школы мы помним, что прямую можно задать уравнением $y = kx + b$
- В таком случае, если наша зависимость (предположительно) описывается прямой, – что нужно сделать, чтобы найти эту оптимальную «лучшую прямую»?

Линейные модели

- С тем, что мы бы хотели сделать с прямой, мы более-менее разобрались. Но давайте поговорим о самой прямой (или, опять же, гиперплоскости в общем случае).
- Из школы мы помним, что прямую можно задать уравнением $y = kx + b$
- В таком случае, если наша зависимость (предположительно) описывается прямой, – что нужно сделать, чтобы найти эту оптимальную «лучшую прямую»?
- Правильно, нужно определить (найти) параметры k и b .

Линейные модели

- В случае же, если у нас не прямая, а гиперплоскость, наше уравнение переписывается в следующем виде:

$$y = \sum_i w_i x_i + w_0$$

Линейные модели

- В случае же, если у нас не прямая, а гиперплоскость, наше уравнение переписывается в следующем виде:

$$y = \sum_i w_i x_i + w_0$$

- или же иначе:

$$y = \langle \hat{w}, x \rangle$$

Линейные модели

- В случае же, если у нас не прямая, а гиперплоскость, наше уравнение переписывается в следующем виде:

$$y = \sum_i w_i x_i + w_0$$

- или же иначе:

$$y = \langle \hat{w}, x \rangle$$

Можно считать, что к вектору x мы добавили единицу в пару к w_0 ;
в таком случае итоговый результат суммы попарных
произведений координат будет в точности равен скалярному
произведению самих векторов!

Линейные модели

- В случае же, если у нас не прямая, а гиперплоскость, наше уравнение переписывается в следующем виде:

$$y = \sum_i w_i x_i + w_0$$

- или же иначе:

$$y = \langle \hat{w}, x \rangle$$

- Что в данных формулах представляют собой x , w и w_0 ? Что это за математические объекты и какой смысл они в себе несут?

Линейные модели

- В случае же, если у нас не прямая, а гиперплоскость, наше уравнение переписывается в следующем виде:

$$y = \sum_i w_i x_i + w_0$$

- или же иначе:

$$y = \langle \hat{w}, x \rangle$$

- Что в данных формулах представляют собой x , w и w_0 ? Что это за математические объекты и какой смысл они в себе несут?
- В данных формулах x – это строка матрицы объектов-признаков, (которую мы определили в самом начале), то есть вектор; w – называют вектором весов регрессии; а w_0 – свободным членом.

Линейные модели



Линейные модели

- Мы обсудили интуицию по которой находится «идеальная линия», но сейчас перед нами встаёт очень важный вопрос.
- Как же определить на практике данные оптимальные веса и коэффициенты прямой/гиперплоскости для регрессии?



Линейные модели

- Ответ: выбирается функция потерь для регрессии (например MSE, или какая-либо из обсуждавшихся нами на прошлой лекции, или же любая другая), а затем для данной выбранной функции потерь уже решается стандартная оптимизационная задача.

Линейные модели

- Ответ: выбирается функция потерь для регрессии (например MSE, или какая-либо из обсуждавшихся нами на прошлой лекции, или же любая другая), а затем для данной выбранной функции потерь уже решается стандартная оптимизационная задача.
- Например, вот так это будет выглядеть для классической функции потерь MSE:

$$\begin{aligned}\min_{\tilde{y}} MSE(\tilde{y}, y) &= \min_{\tilde{y}} \frac{1}{n} \sum_i^n (\tilde{y}_i - y_i)^2 = \\ &= \min_{\hat{w}} \frac{1}{n} \sum_i^n (\langle \hat{w}, x_i \rangle - y_i)^2\end{aligned}$$

Линейные модели

$$\min_{\hat{w}} \frac{1}{n} \sum_i^n (\langle \hat{w}, x_i \rangle - y_i)^2$$

- Что представляет собой вектор весов \hat{w} с точки зрения модели машинного обучения?

Линейные модели

$$\min_{\hat{w}} \frac{1}{n} \sum_i^n (\langle \hat{w}, x_i \rangle - y_i)^2$$

- Что представляет собой вектор весов \hat{w} с точки зрения модели машинного обучения?
- Ответ: это вектор параметров модели.

Линейные модели

$$\min_{\hat{w}} \frac{1}{n} \sum_i^n (\langle \hat{w}, x_i \rangle - y_i)^2$$

- Что представляет собой вектор весов \hat{w} с точки зрения модели машинного обучения?
- Ответ: это вектор параметров модели.
- Как мы можем получить наилучший вектор весов \hat{w} или, другими словами, как решить эту оптимизационную задачу?

Линейные модели

$$\min_{\hat{w}} \frac{1}{n} \sum_i^n (\langle \hat{w}, x_i \rangle - y_i)^2$$

- Что представляет собой вектор весов \hat{w} с точки зрения модели машинного обучения?
- Ответ: это вектор параметров модели.
- Как мы можем получить наилучший вектор весов \hat{w} или, другими словами, как решить эту оптимизационную задачу?
- Ответ из двух слов: ...

Линейные модели

$$\min_{\hat{w}} \frac{1}{n} \sum_i^n (\langle \hat{w}, x_i \rangle - y_i)^2$$

- Что представляет собой вектор весов \hat{w} с точки зрения модели машинного обучения?
- Ответ: это вектор параметров модели.
- Как мы можем получить наилучший вектор весов \hat{w} или, другими словами, как решить эту оптимизационную задачу?
- Ответ из двух слов: градиентный спуск :)

Связь между линейной регрессией и классификацией

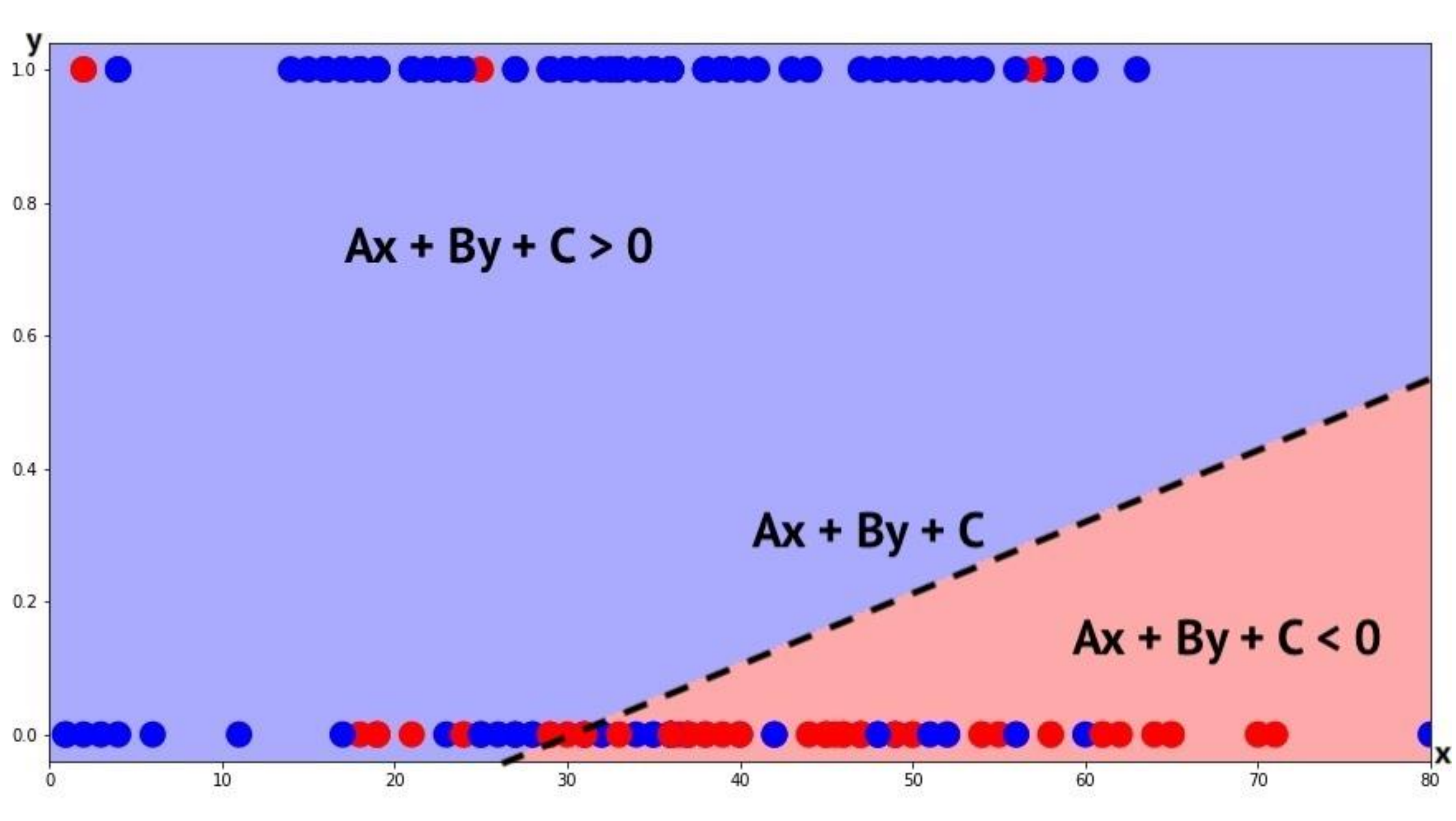
- Всё, что мы сейчас обсуждали, – это прекрасно! Но давайте не будем забывать, что основная тема нашей сегодняшней лекции – это задача классификации. Соответственно, возникает логичный вопрос: а можно ли применить линейные модели для решения задач классификации?

Связь между линейной регрессией и классификацией

- Всё, что мы сейчас обсуждали, – это прекрасно! Но давайте не будем забывать, что основная тема нашей сегодняшней лекции – это задача классификации. Соответственно, возникает логичный вопрос: а можно ли применить линейные модели для решения задач классификации?
- На самом деле, это не так уж просто, в сравнении с задачами регрессии, поскольку в классификации мы пытаемся предсказать значение какой-то категории, а не число, как в регрессии – однако это всё равно возможно!
- Обсудим всё по порядку!

Связь между линейной регрессией и классификацией

- Итак, интуиция: у нас уже есть алгоритм, который умеет неплохо находить линейные зависимости (наша линейная регрессия, которую обсуждали буквально только что).
- Может, мы сможем приспособить этот алгоритм к тому, чтобы нарисовать линию, которая разделит между собой два класса?
- Всё, что здесь нужно, – придумать способ, как обучать его не «ложиться» вдоль точек, а наоборот делить их поперек.



Связь между линейной регрессией и классификацией

- Глядя на предыдущий пример с прямой, нетрудно убедиться, что – при наличии уравнения этой прямой – точки, которые лежат выше прямой, будут в уравнении давать положительное число, а те, которые ниже, – отрицательное.

Связь между линейной регрессией и классификацией

- Глядя на предыдущий пример с прямой, нетрудно убедиться, что – при наличии уравнения этой прямой – точки, которые лежат выше прямой, будут в уравнении давать положительное число, а те, которые ниже, – отрицательное.
- В этом можно убедиться, нарисовав простой пример и подставив точки.
- Аналогично будет работать и в многомерном случае.

Связь между линейной регрессией и классификацией

- Глядя на предыдущий пример с прямой, нетрудно убедиться, что – при наличии уравнения этой прямой – точки, которые лежат выше прямой, будут в уравнении давать положительное число, а те, которые ниже, – отрицательное.
- В этом можно убедиться, нарисовав простой пример и подставив точки.
- Аналогично будет работать и в многомерном случае.
- Как же тогда получить предсказание классификатора?

Связь между линейной регрессией и классификацией

Предсказанием линейного алгоритма для классификации будет являться знак (плюс или минус) выражения, полученного подстановкой координат в уравнение нашей прямой.

Связь между линейной регрессией и классификацией

- Важная ремарка: здесь мы пока везде ведём разговор лишь о бинарной классификации... кстати, что это?

Связь между линейной регрессией и классификацией

- Важная ремарка: здесь мы пока везде ведём разговор лишь о бинарной классификации, то есть классификации объектов между двумя классами (категориями).

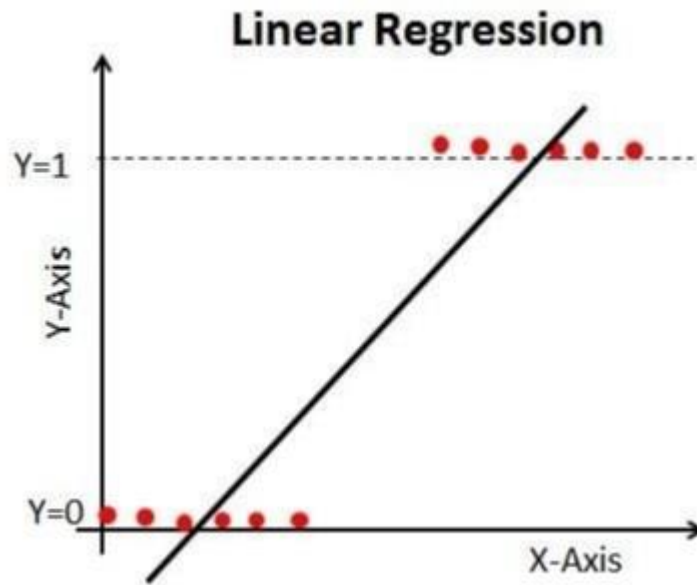
Связь между линейной регрессией и классификацией

- Важная ремарка: здесь мы пока везде ведём разговор лишь о бинарной классификации, то есть классификации объектов между двумя классами (категориями).
- Если же классов у нас в данных больше двух (так называемая многоклассовая классификация), – в таком случае применяются уже немного более сложные подходы, которые мы обязательно обсудим на семинаре!

Связь между линейной регрессией и классификацией

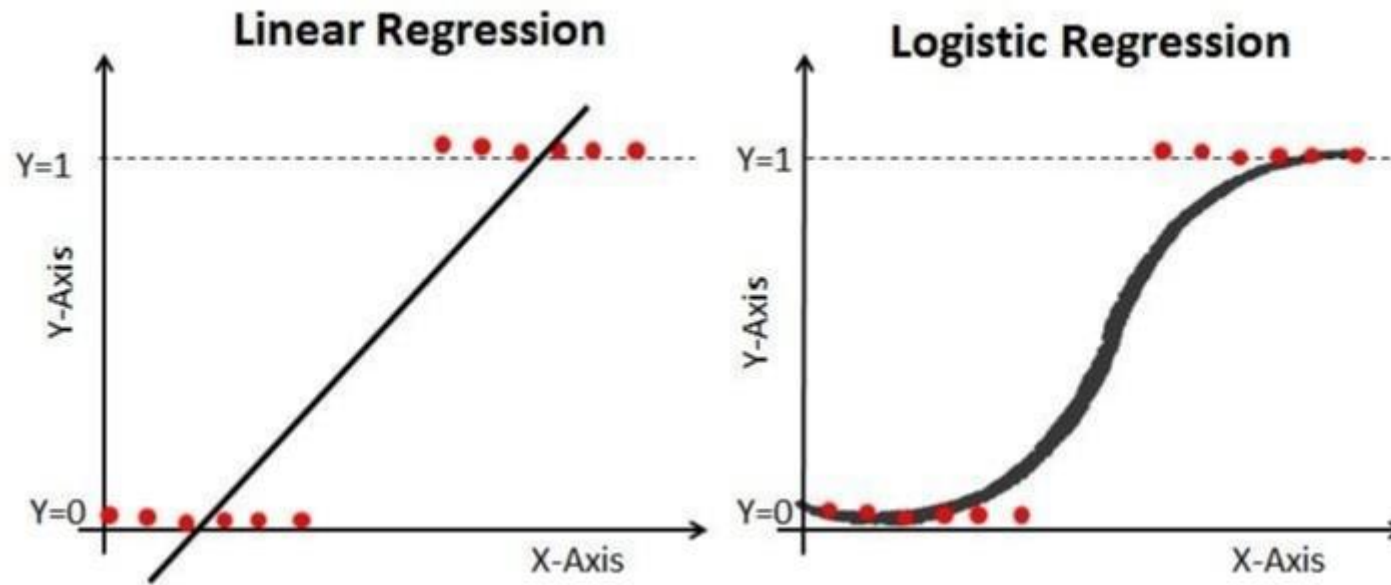
- Отлично, мы с вами разобрались в целом с графической интуицией задачи.
- Но давайте теперь посмотрим на эту задачу не в разрезе признак/признак, а в разрезе признак/целевая переменная.
- Какие проблемы возникают в этом случае?

Связь между линейной регрессией и классификацией



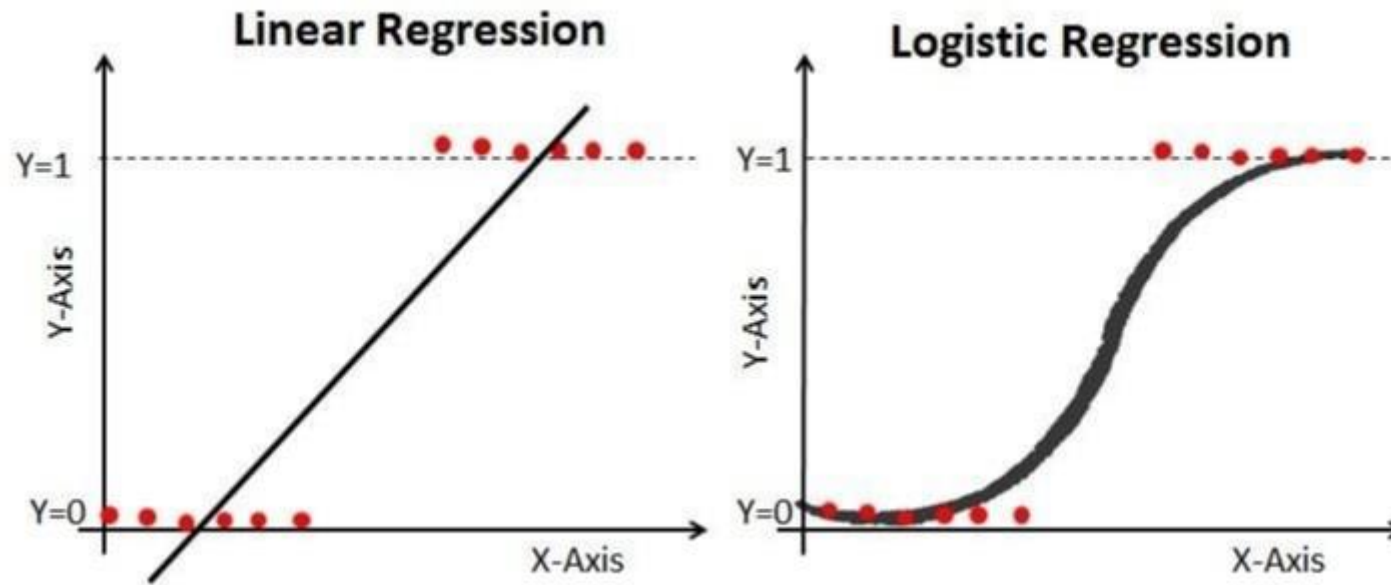
- Проблема в том, что y нас по смыслу не может и не должно быть значений выше $Y = 1$ и ниже $Y = 0$.
- Значит, нам нужно как-то преобразовать наше линейное отображение.

Связь между линейной регрессией и классификацией



- По сути, нам нужно из левой картинке сделать правую, применив какое-нибудь преобразование, и тем самым мы сможем обучать обычную линейную регрессию, которая будет пытаться аппроксимировать классы 0 и 1.

Связь между линейной регрессией и классификацией

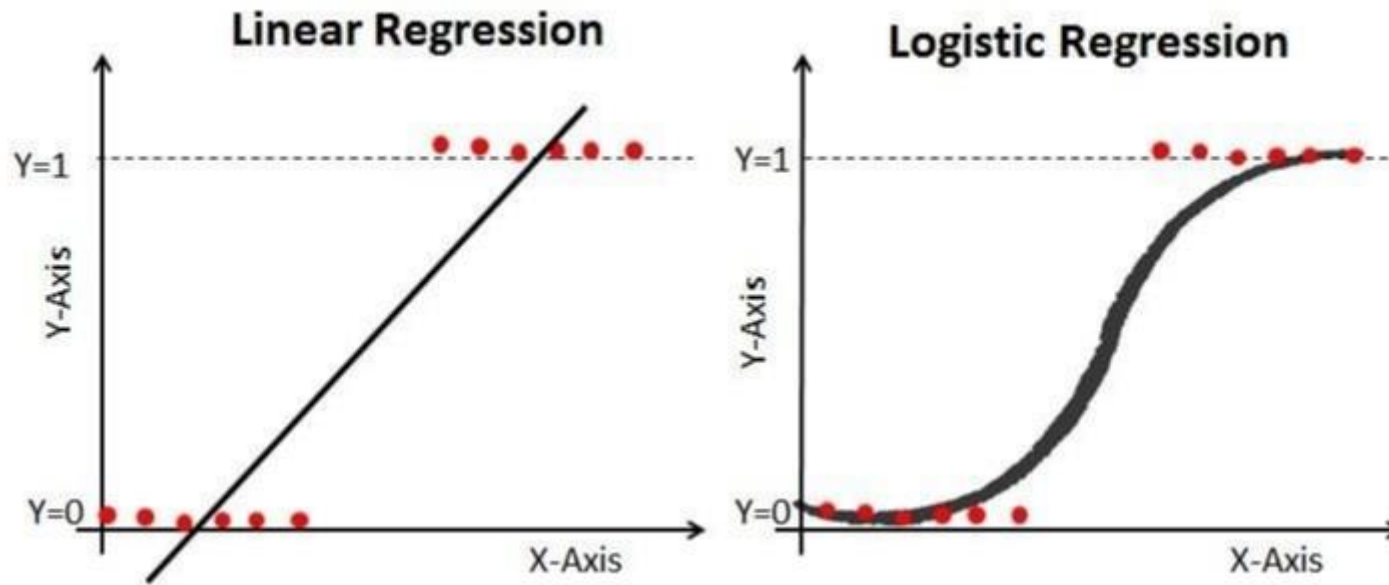


Вопрос на понимание:
а как, имея правую
кривую, всё же
перейти к итоговому
предсказанию
классификатора?

- По сути, нам нужно из левой картинки сделать правую, применив какое-нибудь преобразование, и тем самым мы сможем обучать обычную линейную регрессию, которая будет пытаться аппроксимировать классы 0 и 1.

Связь между линейной регрессией и классификацией

Подробнее о том, как логистическая регрессия реализуется с математической и технической точки зрения, мы с вами поговорим на семинарах!



- По сути, нам нужно из левой картинке сделать правую, применив какое-нибудь преобразование, и тем самым мы сможем обучать обычную линейную регрессию, которая будет пытаться аппроксимировать классы 0 и 1.