Лекция 6
«Логистическая регрессия»

Автор: д.т.н Джомартова Ш.А.

Как работает линейная регрессия?

1. Fit - поиск оптимальных весов



2. Predict - использование весов для предсказания

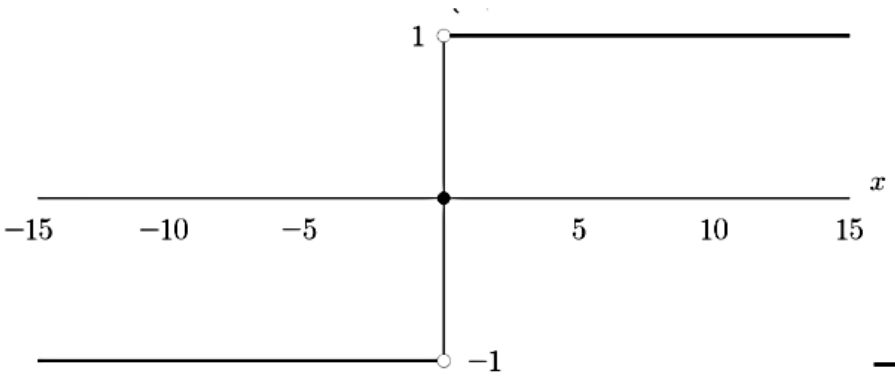


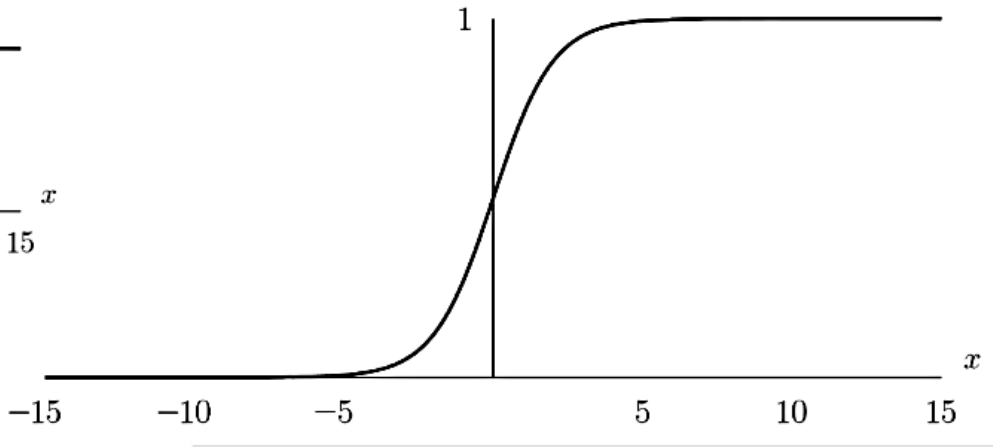
Чем будет отличаться линейная классификация?

1. **Будем более хитро комбинировать веса и признаки между собой**

$$sign(z) = \begin{cases} 1 & , z > 0 \\ 0 & , z = 0 \\ -1 & , z < 0 \end{cases}$$

$$\sigma(z) = \frac{1}{1 + \exp(-z)}$$

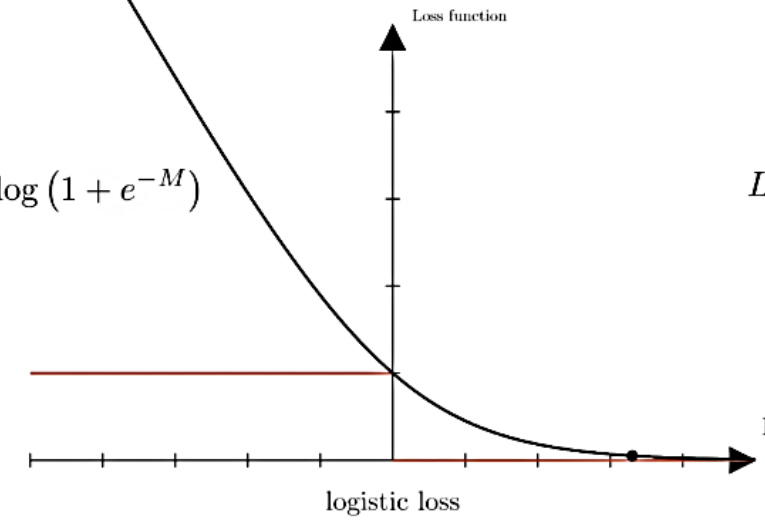




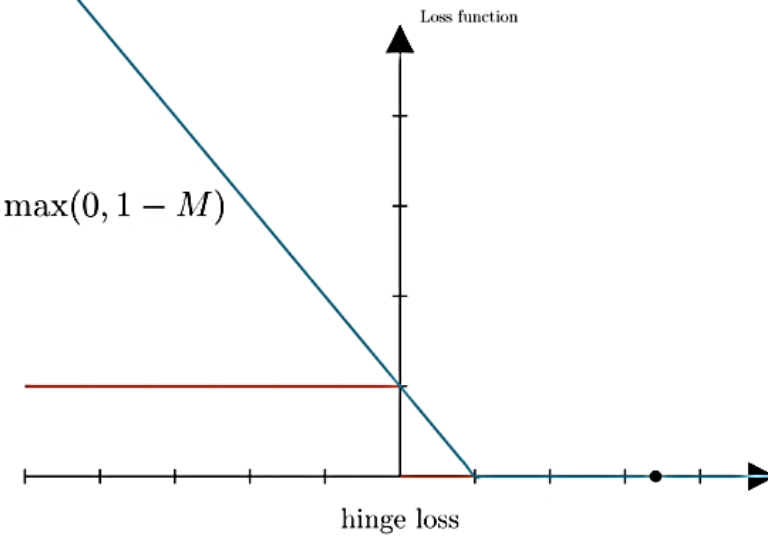
Чем будет отличаться линейная классификация?

2. Будем использовать другие функционалы потерь

$L(M) = \log\left(1 + e^{-M}\right)$



$L(M) = \max(0, 1 - M)$



Модельная задача: выдача кредитов

Дано: характеристики заёмщиков, их кредитная история

Задача: определить, кому выдавать кредит, а кому нет



Машина должна говорить либо да, либо нет

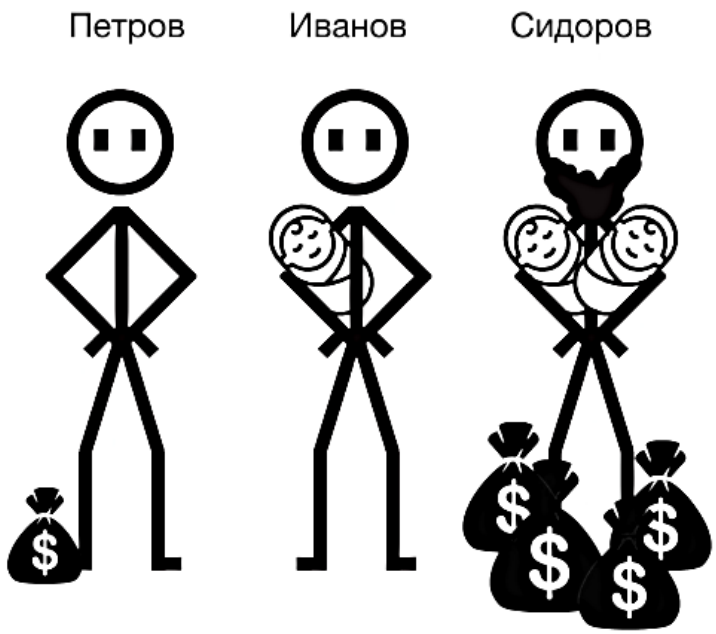
Пусть положительный класс - это
"да", отрицательный - это "нет"



Пример: к нам пришли заёмщики

Фичи (характеристики)

	x_1	x_2	x_3
Заемщики	Зарплата	Возраст	Количество детей
Иванов	0	18	1
Петров	20 000	18	0
Сидоров	150 000	30	2



Предобработаем данные

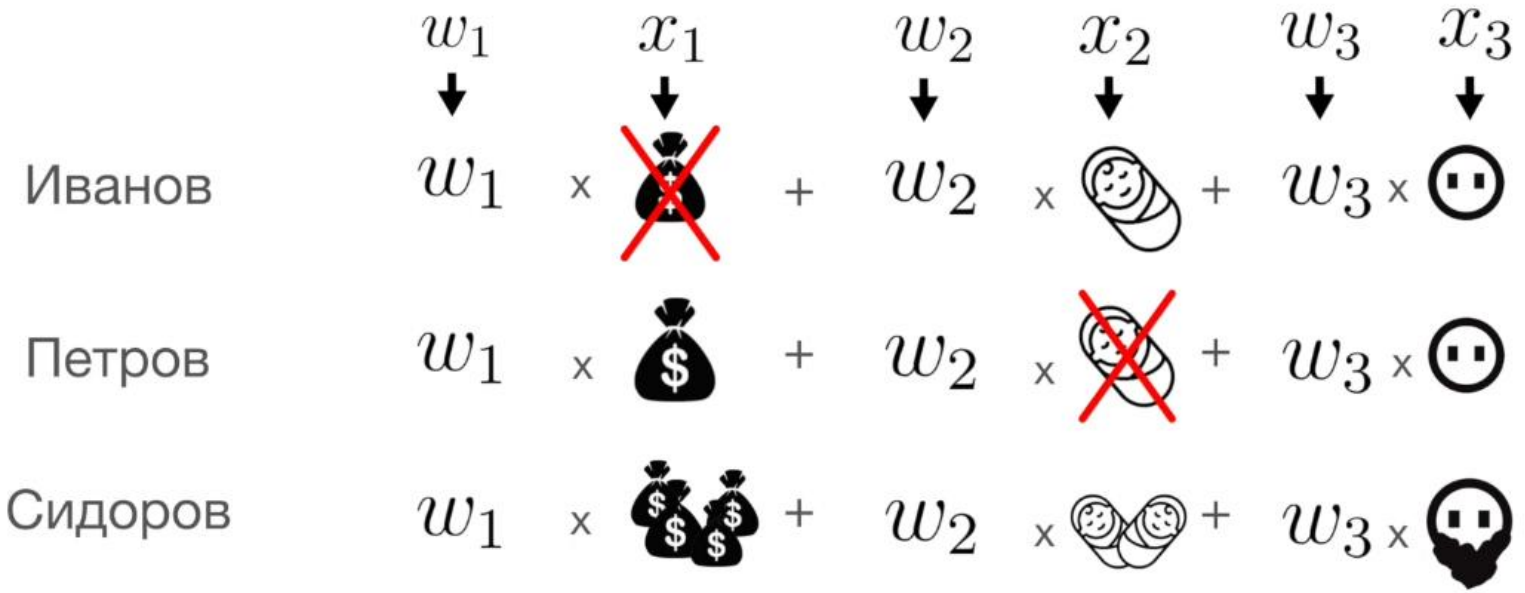
	x_1	x_2	x_3
	Зарплата	Возраст	Количество детей
Иванов	0	18	1
Петров	20 000	18	0
Сидоров	150 000	30	2



MinMaxScaler

	x_1	x_2	x_3
	Зарплата	Возраст	Количество детей
Иванов	0	0.3	0.2
Петров	0.2	0.3	0
Сидоров	0.6	0.7	0.4

Обучить линейную модель - значит найти оптимальные веса



$$\vec{w} = (329.5, -285.5, -143)$$

	w_1	x_1	w_2	x_2	w_3	x_3
Иванов	$329.5 \times$		$+ -285.5 \times$		$+ -143 \times$	
Петров	$329.5 \times$		$+ -285.5 \times$		$+ -143 \times$	
Сидоров	$329.5 \times$		$+ -285.5 \times$		$+ -143 \times$	

Вспомним, что такое модель линейной регрессии - это **произведение вектора весов на вектор признаков**

$$a(x) = \langle w, x \rangle$$

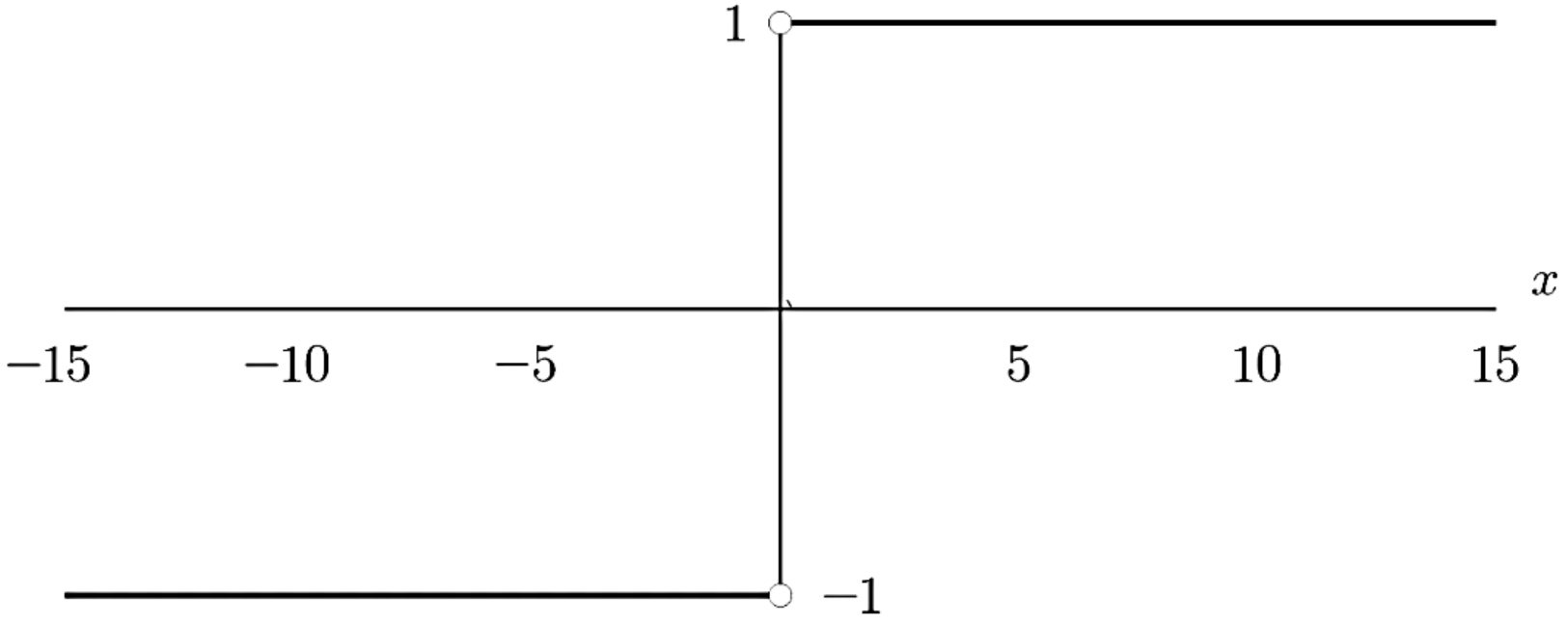
его мы уже нашли

Выдача модели - вещественное число

	w_1	x_1		w_2	x_2		w_3	x_3		
Иванов	329.5	x 0	+	-285.5	x 0.2	+	-143	x 0.3	=	-100
Петров	329.5	x 0.2	+	-285.5	x 0	+	-143	x 0.3	=	23
Сидоров	329.5	x 0.6	+	-285.5	x 0.4	+	-143	x 0.5	=	12

Воспользуемся функцией $sign$

$$sign(x) = \begin{cases} 1 & , x > 0 \\ 0 & , x = 0 \\ -1 & , x < 0. \end{cases}$$

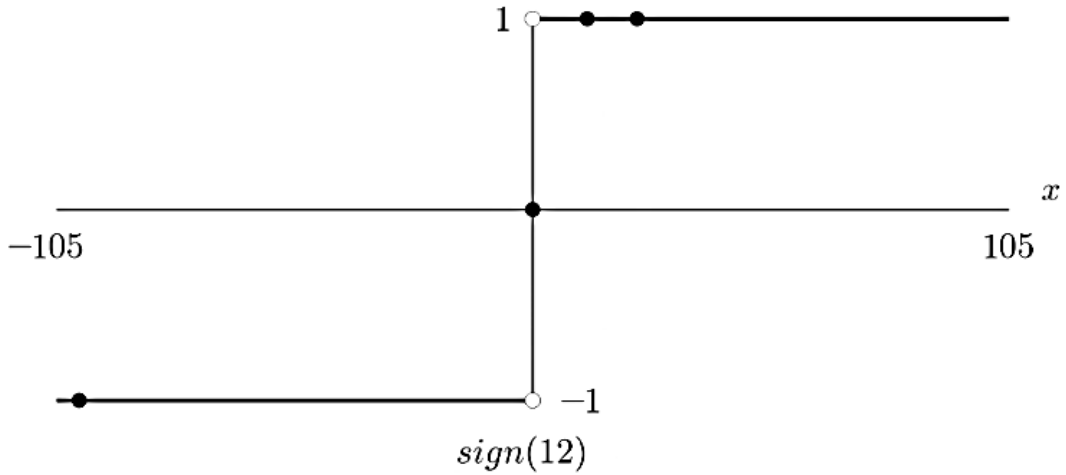


Классифицируем по знаку

Заемщики	Выдача линейной регрессии
Иванов	-100
Петров	23
Сидоров	12

Выдача линейной классификации
-1
+1
+1

$$sign(x) = \begin{cases} 1 & , x > 0 \\ 0 & , x = 0 \\ -1 & , x < 0. \end{cases}$$



Заемщики	Выдача линейной регрессии
Иванов	-100
Петров	23
Сидоров	12

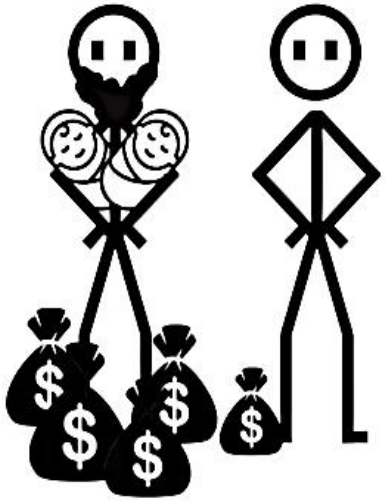
Выдача линейной классификации
-1
+1
+1



Петров Сидоров



Иванов





1. Предобрабатываем данные

	x_1 Зарплата	x_2 Возраст	x_3 Количество детей
Иванов	0	18	1
Петров	20 000	18	0
Сидоров	150 000	30	2

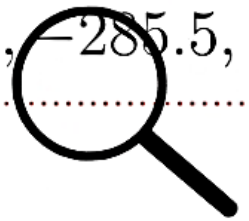


MinMaxScaler

	x_1 Зарплата	x_2 Возраст	x_3 Количество детей
Иванов	0	0.3	0.2
Петров	0.2	0.3	0
Сидоров	0.6	0.7	0.4

2. Ищем веса

$$\vec{w} = (329.5, -285.5, -143)$$



3. Используем $sign$

Заемщики	Выдача линейной регрессии
Иванов	-100
Петров	23
Сидоров	12



Заемщики	Выдача линейной классификации
Иванов	-1
Петров	+1
Сидоров	+1

Получилась модель:

$$a(x) = sign(\langle w, x \rangle)$$

Чтобы модель училась лучше, надо говорить, что хорошо, а что плохо





Для этого нужны функционалы потерь

Посчитаем долю ошибок

Модель 1



a y

Заемщики	Прогноз	Факт	Наличие ошибки
Иванов	-1	+1	1
Петров	+1	+1	0
Сидоров	+1	-1	1

$$\frac{1+0+1}{1+1+1} = 0.666...$$
– количество наблюдений

Модель 2



a y

Заемщики	Прогноз	Факт	Наличие ошибки
Иванов	+1	+1	0
Петров	+1	+1	0
Сидоров	+1	-1	1

$$\frac{0+0+1}{1+1+1} = 0.333...$$
– количество наблюдений

Получается, наша функция потерь будет выглядеть так:

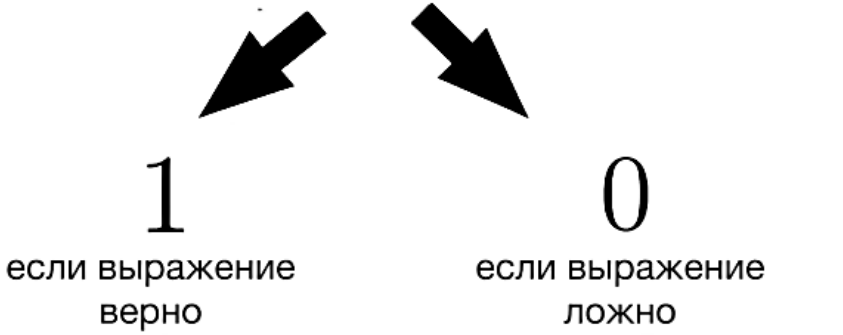
$$Q(a, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) \neq y_i]$$

l - количество наблюдений

$a(x_i)$ - предсказания алгоритма

y_i - фактические значения

$[a(x_i) \neq y_i]$ - индикатор



Но для модели: $a(x) = sign(\langle w, x \rangle)$

-10 **и** $-10\ 000$ **эквивалентны**

Заемщики	Выдача линейной регрессии
Иванов	-10
Петров	-10000



Заемщики	Выдача линейной классификации
Иванов	-1
Петров	-1

Рассмотрим вот такое выражение

$$M_i = y_i \langle w, x_i \rangle$$

По нему мы можем определить не только **корректность** ответа, но и **уверенность** в нем



$\langle w, x \rangle$	y	$Margin$
Скалярное произведение	Факт	Отступ
0.2	+1	0.2

$$M_i = +1 \times 0.2 > 0$$



$\langle w, x \rangle$	y	$Margin$
Скалярное произведение	Факт	Отступ
-0.1	+1	-0.1

$$M_i = +1 \times (-0.1) < 0$$

Эта величина называется margin (отступ), с помощью нее можно переписать текущий функционал потерь

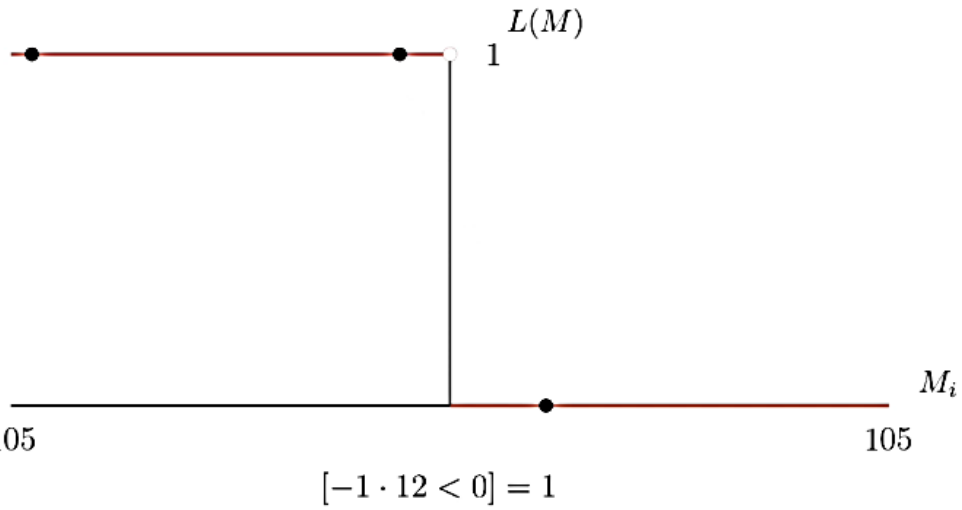
$$Q(a, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) \neq y_i]$$

$$Q(a, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} [\underbrace{y_i \langle w, x_i \rangle}_{M_i} < 0]$$

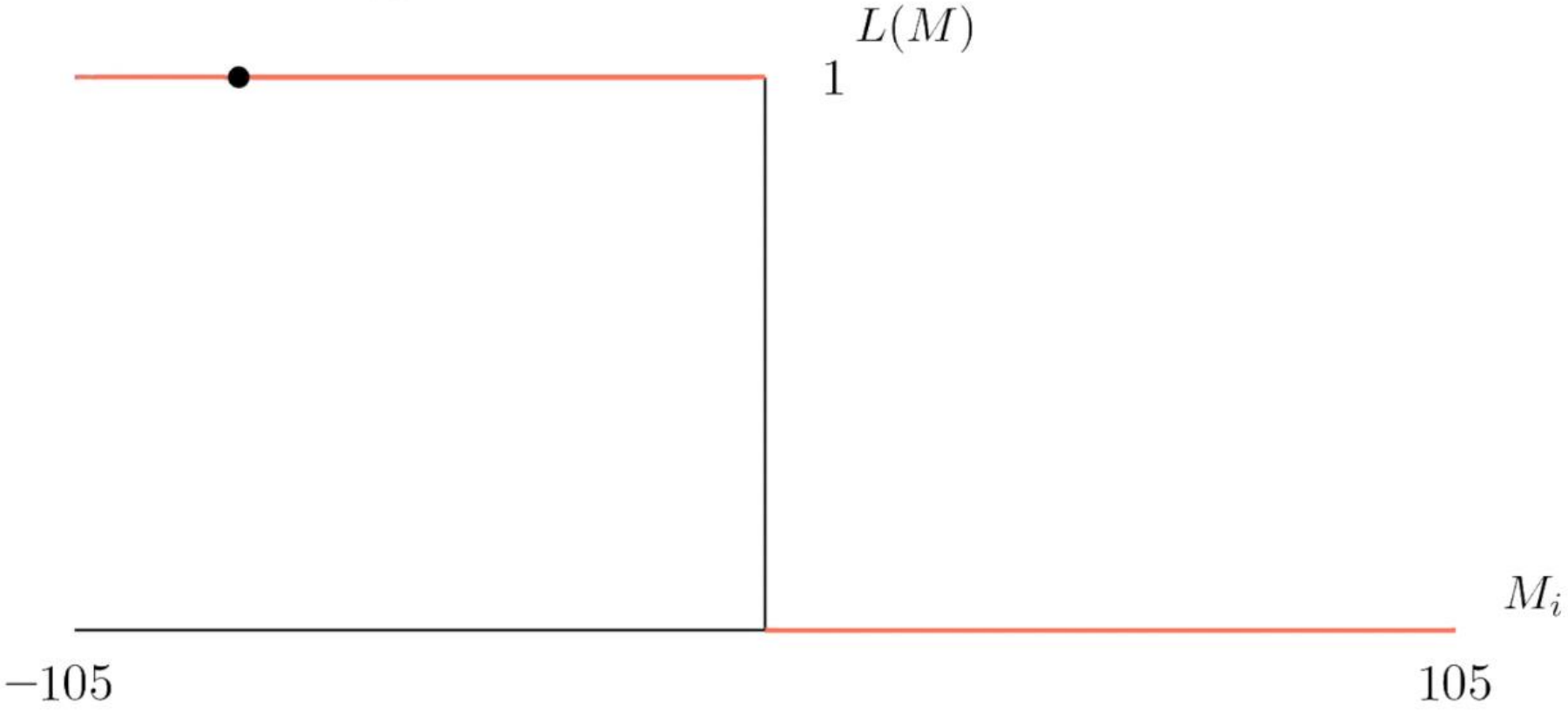
Посмотрим на график нашей функции потерь

	$\langle w, x \rangle$	y	$Margin$
Заемщики	Скалярное произведение	Факт	Отступ
Иванов	-100	+1	-100
Петров	23	+1	23
Сидоров	12	-1	-12

$$L(M) = [\underbrace{y_i \langle w, x_i \rangle}_{M_i} < 0]$$

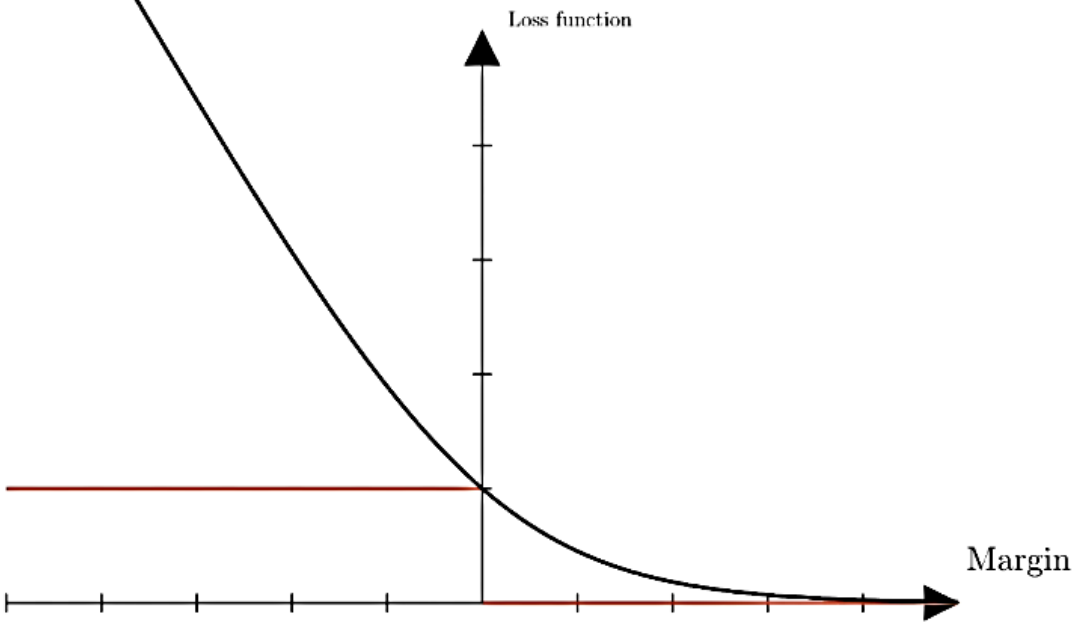


Это не гладкая функция, а градиентный спуск работает только на гладких



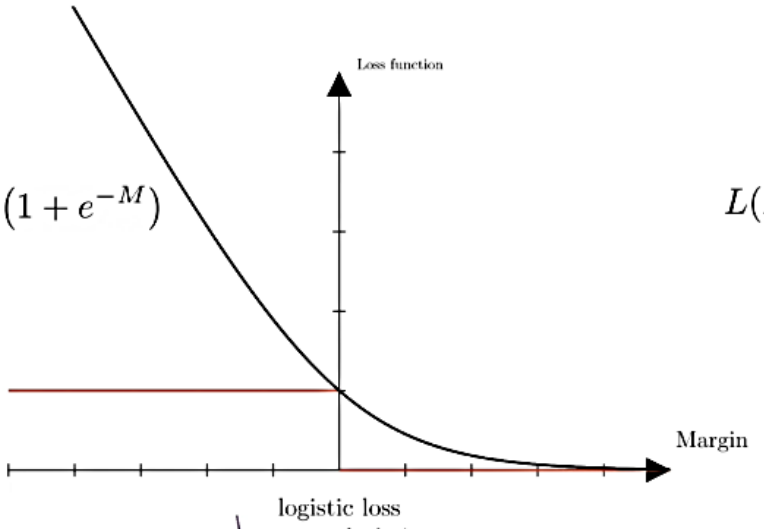
Допустим, это будет **логистическая** функция потерь:

$$L(M) = \log\left(1 + e^{-M}\right)$$



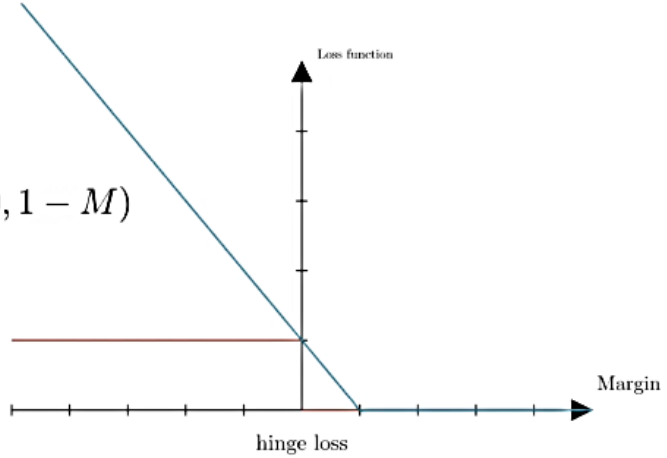
logistic loss

$L(M) = \log\left(1 + e^{-M}\right)$



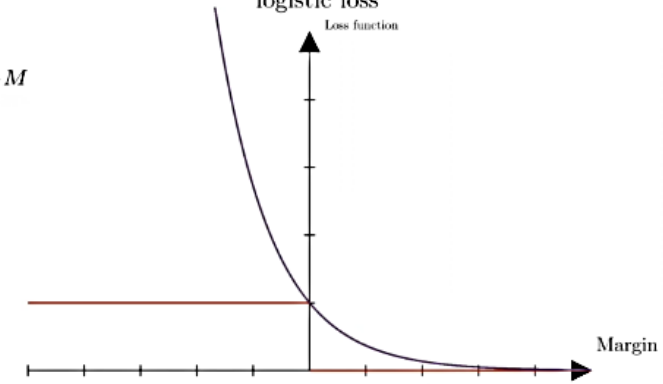
logistic loss

$L(M) = \max(0, 1 - M)$



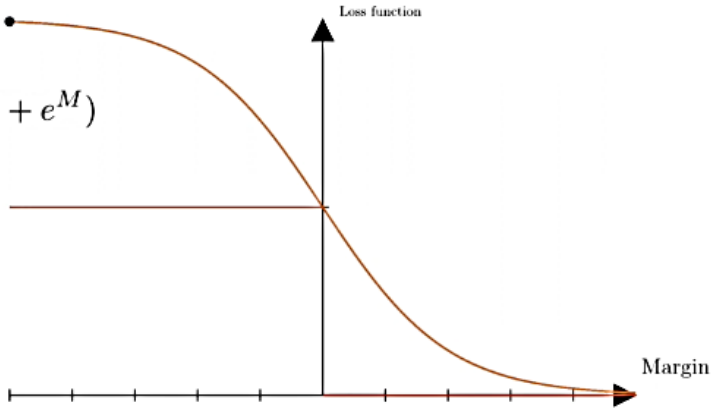
hinge loss

$L(M) = e^{-M}$



exponential loss

$L(M) = 2/(1 + e^{M})$



sigmoid loss



Как выдавать не класс, а вероятность класса?

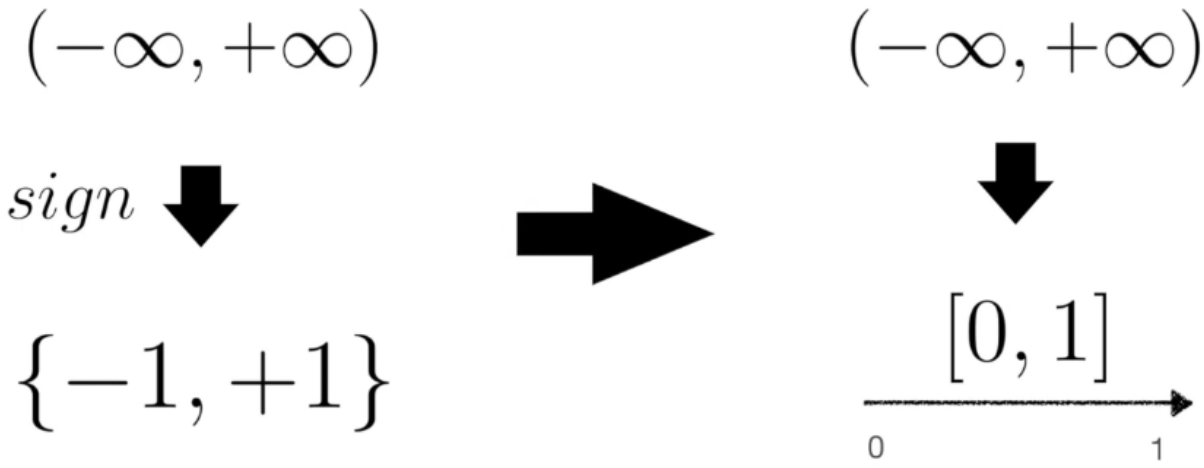
Мы хотим посчитать **вероятность** того, что клиент **вернёт кредит**

Если **вероятность ниже 80%** - будем отправлять заявку на дополнительную проверку

Но откуда нам взять вероятность?



Если наша модель выдает либо да, либо нет, мы не понимаем, насколько она уверена в ответе



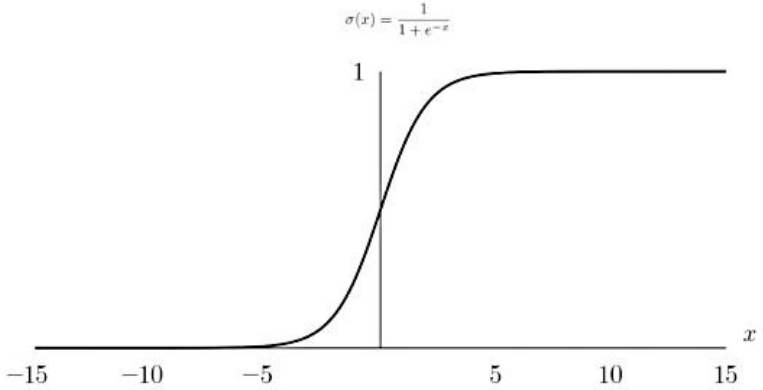
Сменим пространство ответов

Во-первых, нам нужно перевести значение скалярного произведения с диапазона $(-\infty, +\infty)$ **на** $[0, 1]$

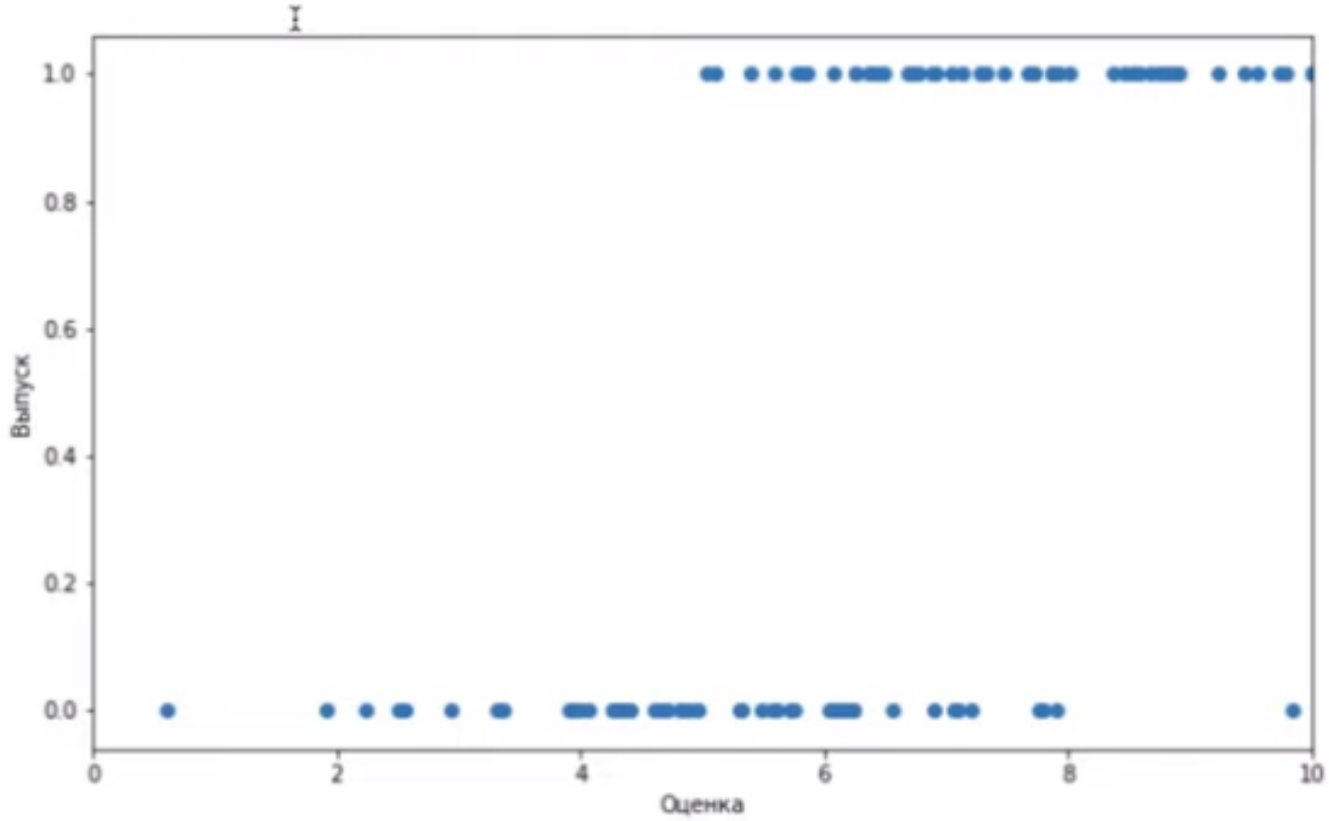
Одна из самых популярных функций для этого - сигмоида



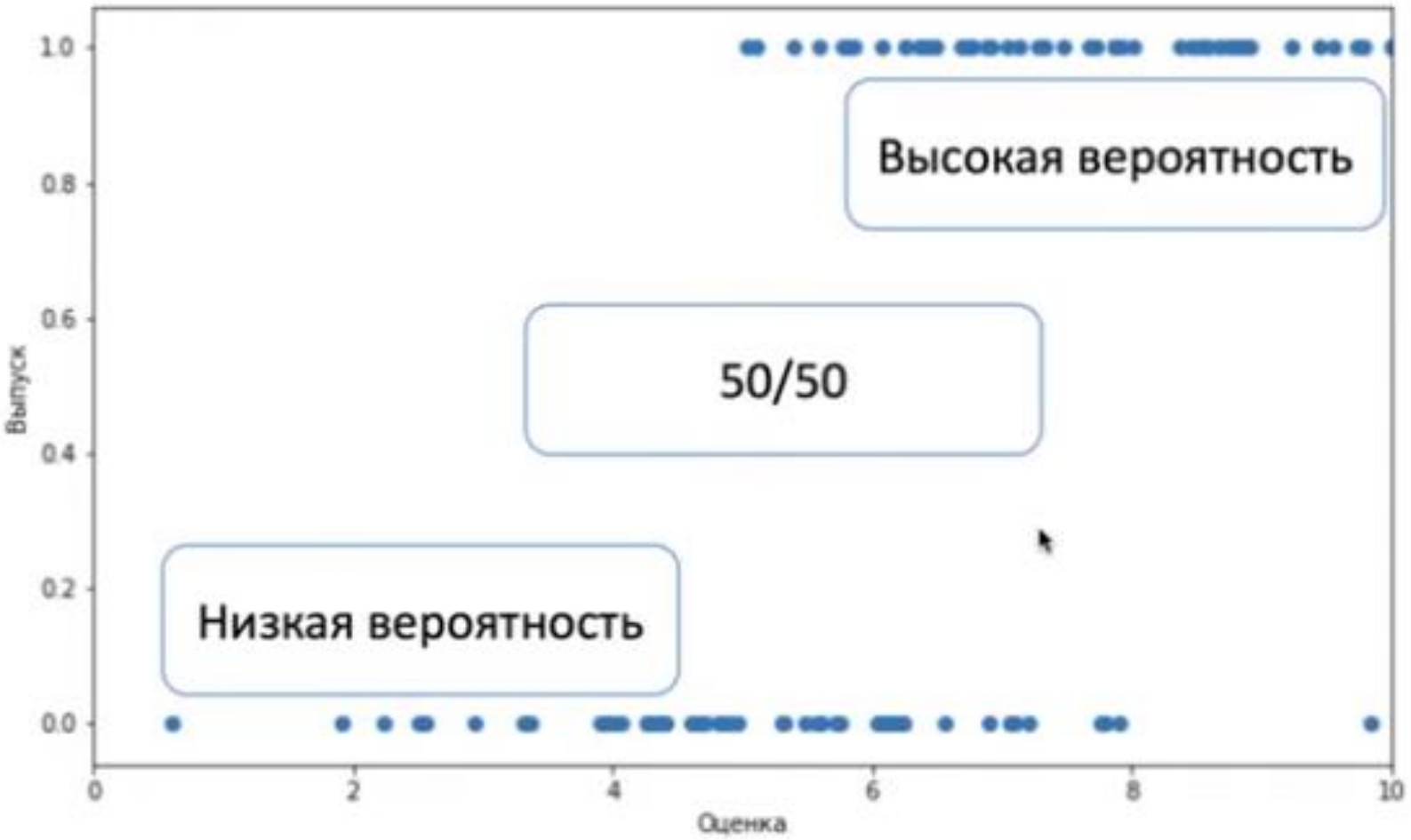
$$\sigma(z) = \frac{1}{1 + \exp(-z)}$$



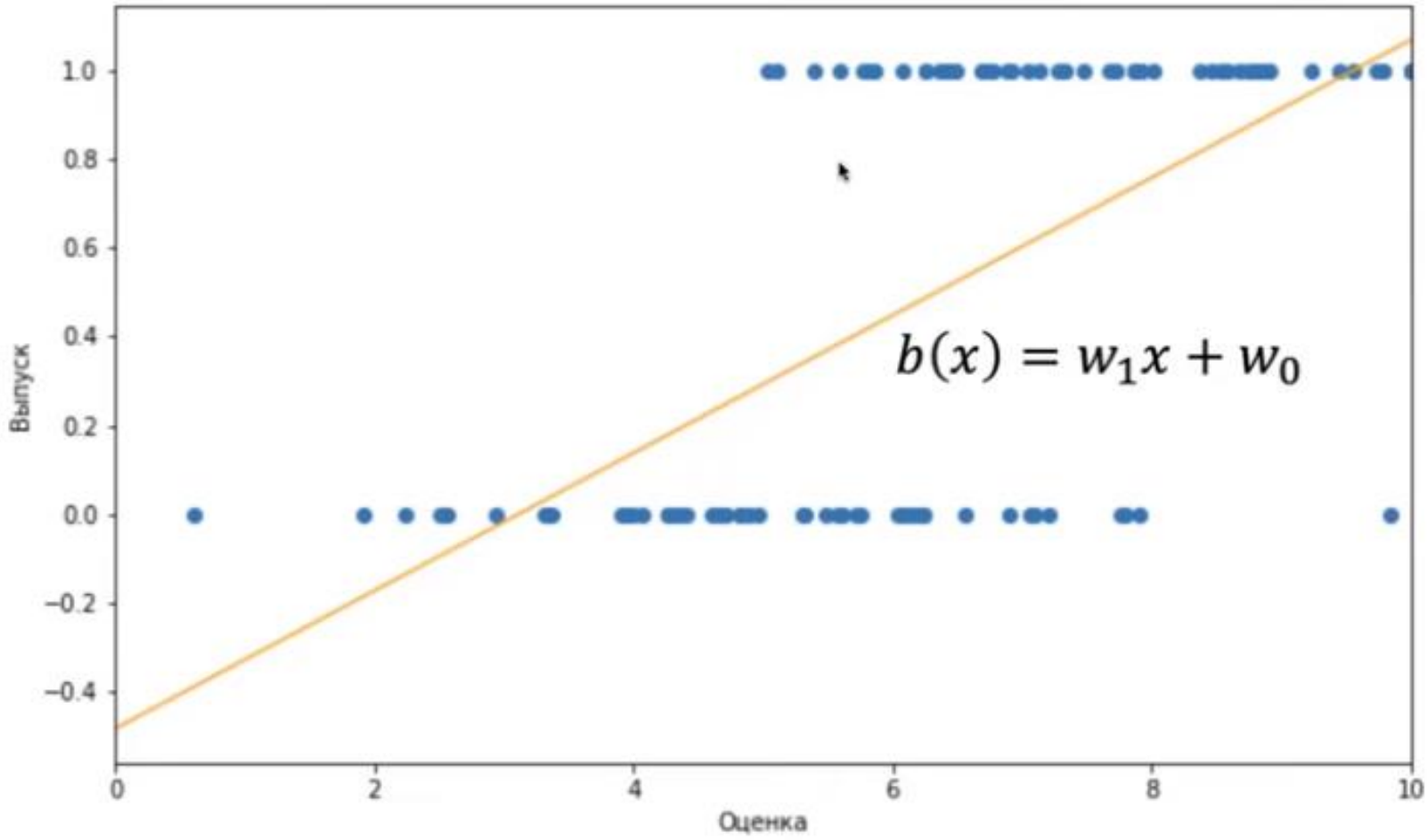
Предсказание вероятностей



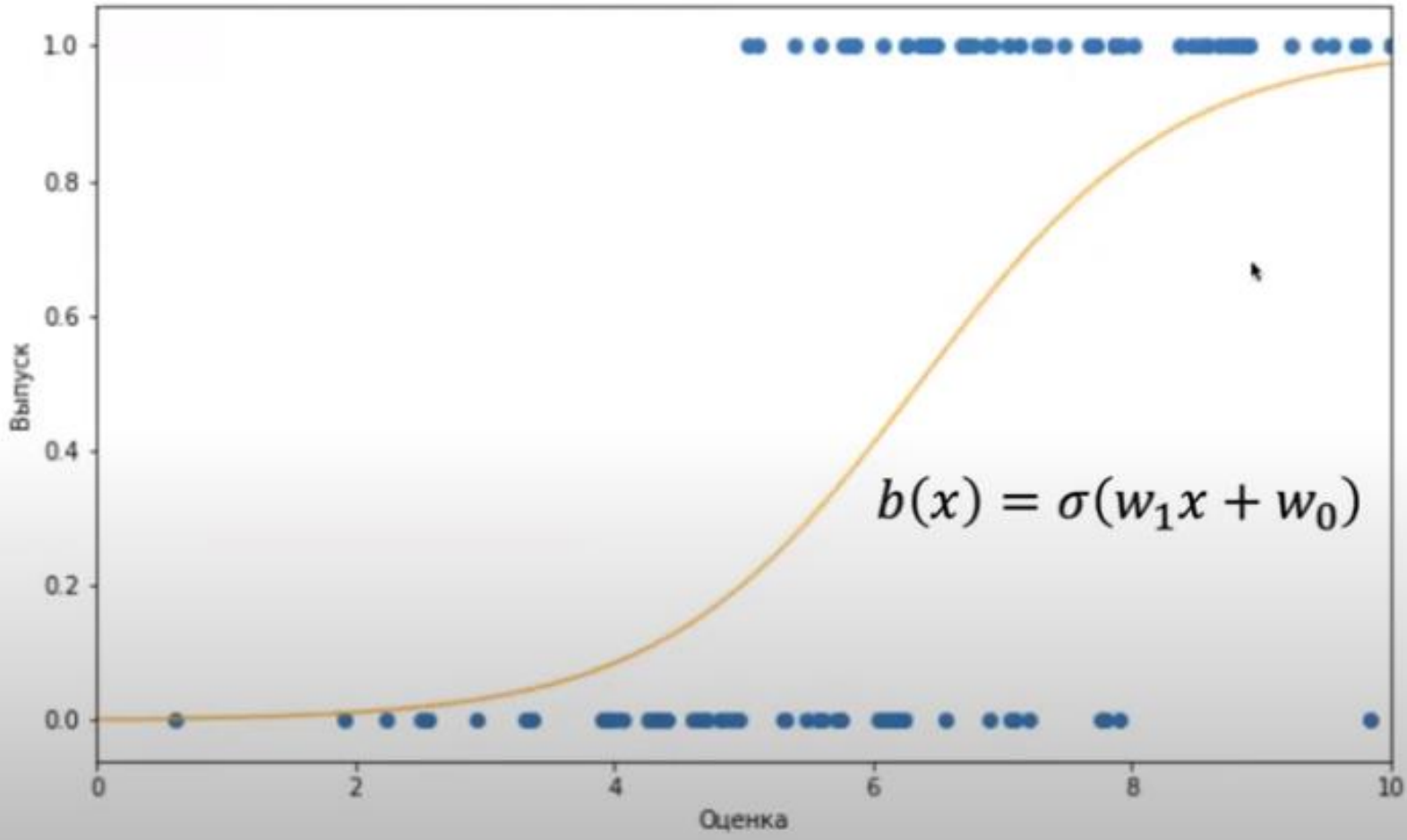
Предсказание вероятностей



Предсказание вероятностей



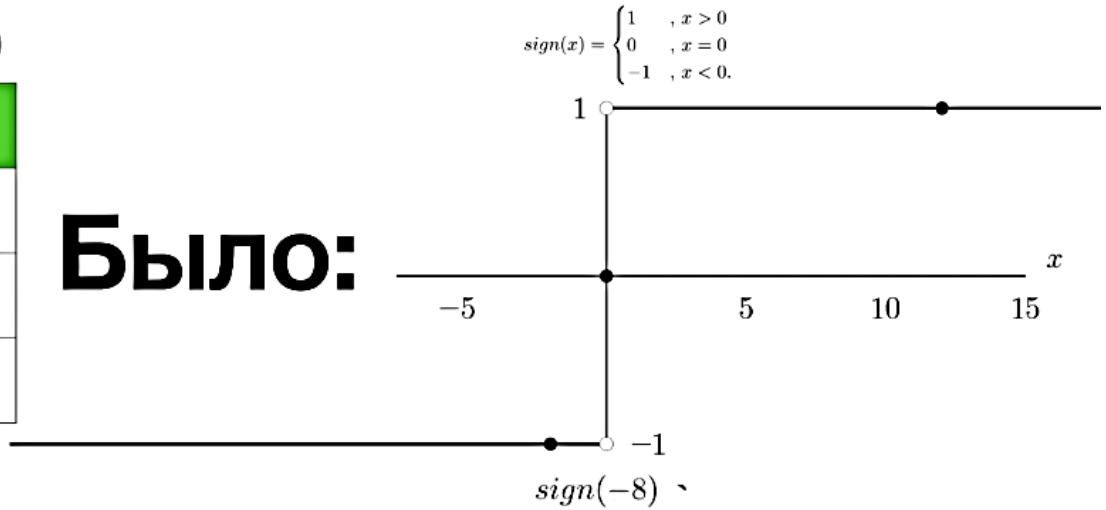
Предсказание вероятностей



$\langle w, x \rangle$ $sign(\langle w, x \rangle)$

Заемщики	Скалярное произведение	Сигмоида
Иванов	-2	-1
Петров	12	1
Сидоров	-8	-1

Было:

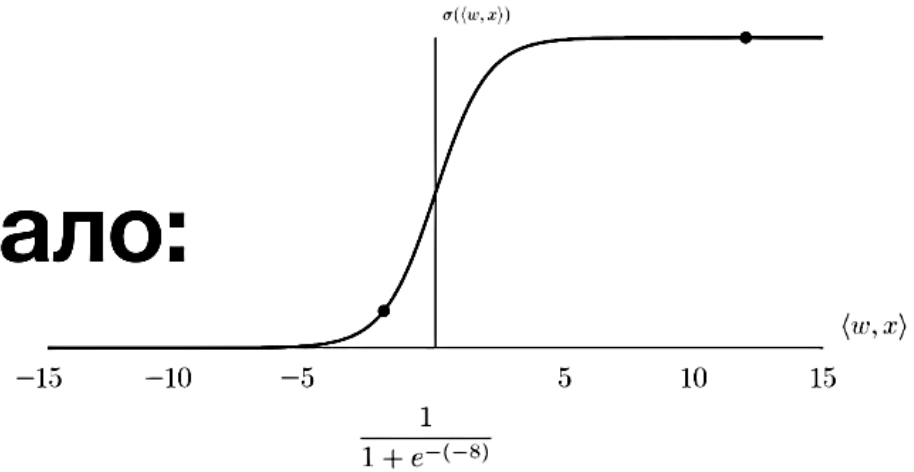


$sign(-8)$,

$\langle w, x \rangle$ $\sigma(\langle w, x \rangle)$

Заемщики	Скалярное произведение	Сигмоида
Иванов	-2	0.119...
Петров	12	0.999993...
Сидоров	-8	0.0003...

Стало:



Первый пункт выполнен: мы перенесли значение скалярного произведения с диапазона $(-\infty, +\infty)$ **на** $[0, 1]$



Второй пункт: нам нужно, чтобы ответ отображал **вероятность**

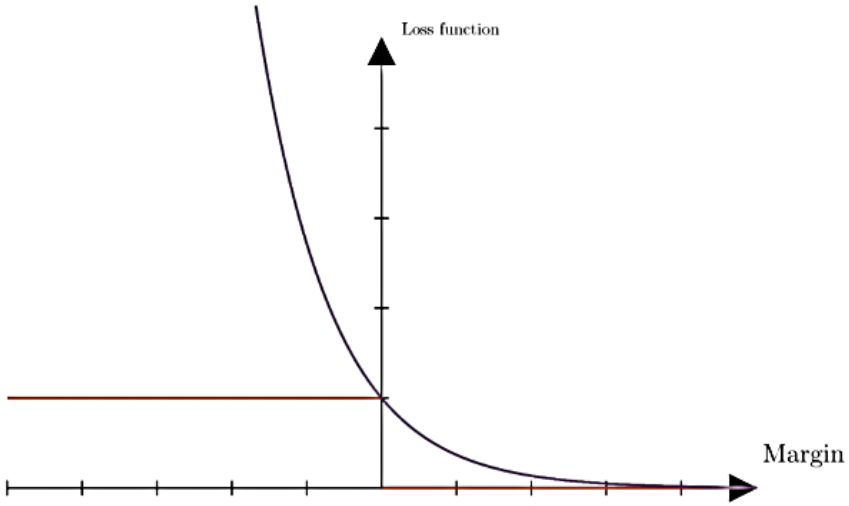
Допустим, модель выдает нам вероятность положительного класса 0.8. Что это значит?



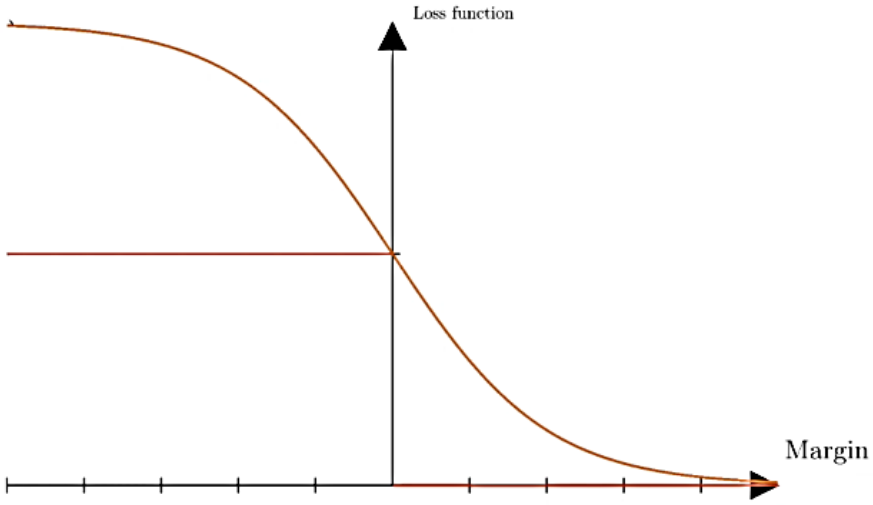
Это значит, что в 10 объектах с такой вероятностью 8 из них имеют положительный класс

Что мы можем использовать, чтобы управлять ответами модели?

Правильно, функционалы потерь



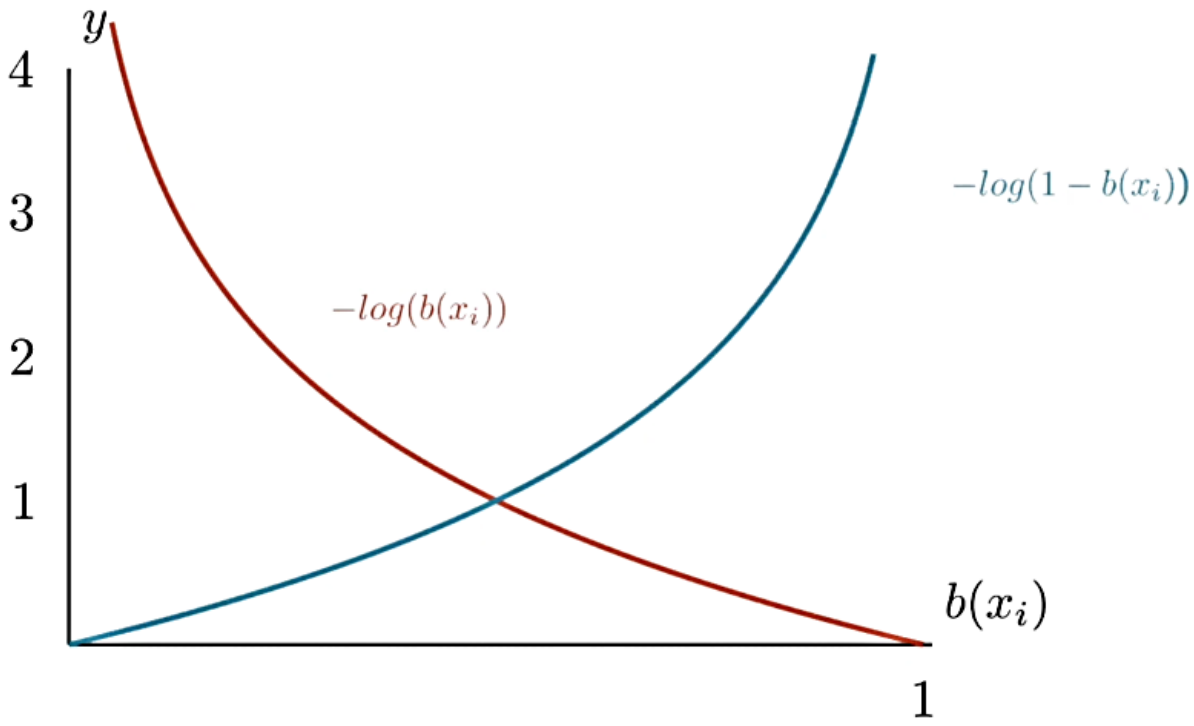
exponential loss



sigmoid loss

объект принадлежит к классу 1 объект принадлежит к классу 0

$$logloss = -\sum_{i=1}^{l} [y_i = +1] \log\left(b(x_i)\right) + [y_i = -1] \log\left(1 - b(x_i)\right)$$



Нам нужно увеличить вероятность правильной классификации

$b(x_i)$ - **вероятность класса** $+1$

Рассмотрим выражение $b(x_i)^{[y_i=+1]}(1-b(x_i))^{[y_i=-1]}$

вероятность правильной классификации, если класс +1

вероятность правильной классификации, если класс -1

Запишем правдоподобие выборки

$$Q(a,X)=\prod_{i=1}^{\ell} b(x_i)^{[y_i=+1]}(1-b(x_i))^{[y_i=-1]}$$

У нас может быть очень много объектов, тогда это произведение будет очень маленьким. Чтобы исправить это, логарифмируем произведение

$$Q(a, X) = \prod_{i=1}^{\ell} b(x_i)^{[y_i=+1]}(1 - b(x_i))^{[y_i=-1]}.$$

Нам удобнее минимизировать функционал, поэтому минус

Логарифмирование - монотонное преобразование, поэтому выполняем его

$$-\sum_{i=1}^{\ell} ([y_i = +1] \log b(x_i) + [y_i = -1] \log(1 - b(x_i)))$$

Подставим сигмоиду в LogLoss

$$\sigma(\langle w, x\rangle) \rightarrow LogLoss(\sigma(\langle w, x\rangle))$$

$$-\sum_{i=1}^{\ell}\left([y_i=+1]\log\frac{1}{1+\exp(-\langle w,x_i\rangle)}+[y_i=-1]\log\frac{\exp(-\langle w,x_i\rangle)}{1+\exp(-\langle w,x_i\rangle)}\right)=$$

$$=-\sum_{i=1}^{\ell}\left([y_i=+1]\log\frac{1}{1+\exp(-\langle w,x_i\rangle)}+[y_i=-1]\log\frac{1}{1+\exp(\langle w,x_i\rangle)}\right)=$$

$$=\sum_{i=1}^{\ell}\log\left(1+\exp(-y_i\langle w,x_i\rangle)\right)$$

$$\sigma(z)=\frac{1}{1+exp(-z)}$$

$$LogLoss=-\sum_{i=1}^{\ell}(y_i=+1]\log b(x_i)+[y_i=-1]\log(1-b(x_i)))$$

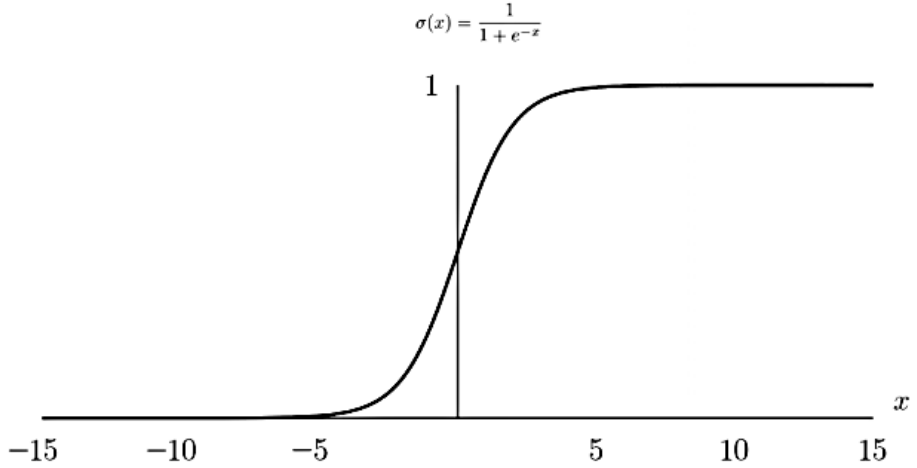
сложнослайд

У нас получился логистический функционал потерь!

$$\sum_{i=1}^{\ell} \log\left(1 + \exp(-y_i \langle w, x_i \rangle)\right)$$

1. Итак, регрессию можно использовать для решения задач классификации, для этого на выход накидывается сигмоида

$$\sigma(z) = \frac{1}{1 + \exp(-z)}$$



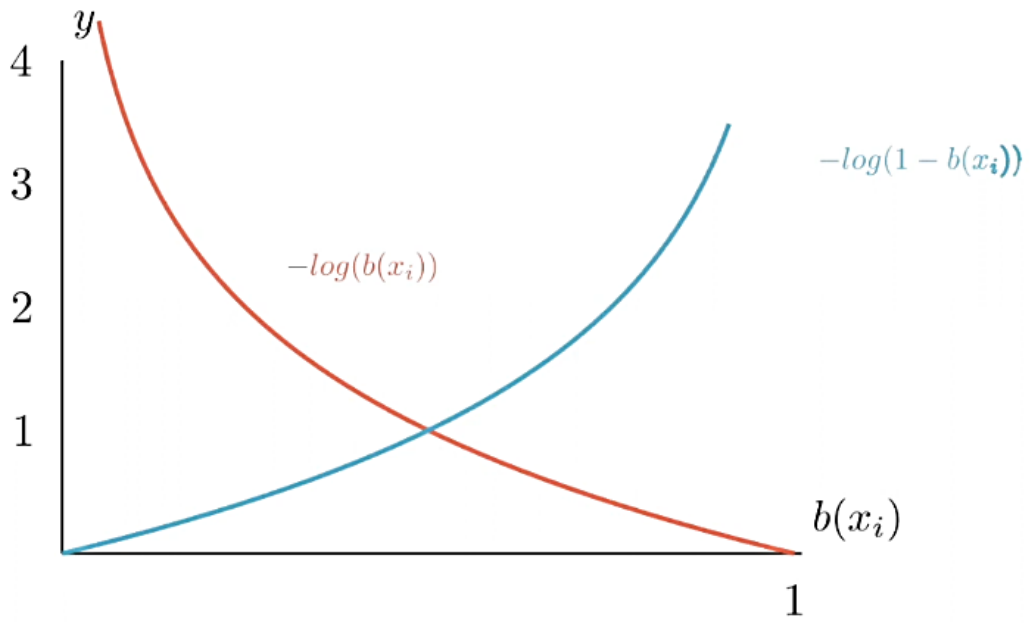
$$(-\infty, +\infty)$$

$$[0, 1]$$

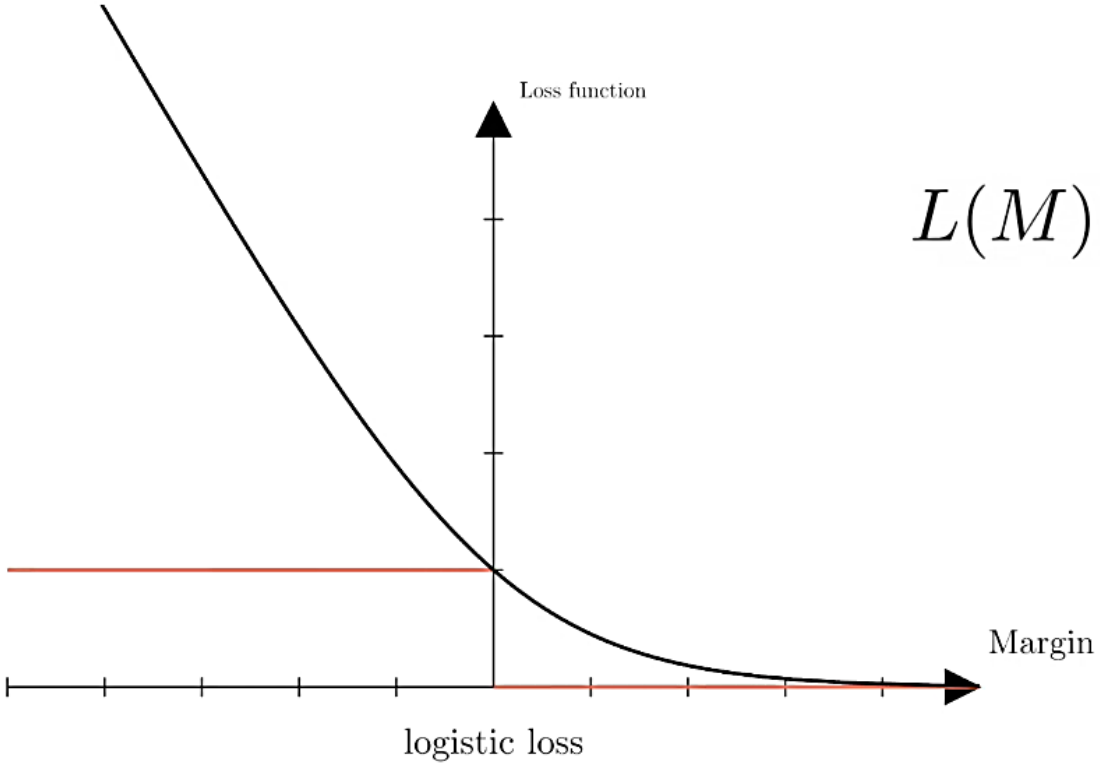
0 1

2. LogLoss - функционал, который грамотно апроксимирует вероятности

$$logloss = -\sum_{i=1}^{l} [y_i = +1] \log (b(x_i)) + [y_i = -1] \log (1 - b(x_i))$$



3. Если мы их сгруппируем, то получится **логистическая функция потерь**



$$L(M) = \log\left(1 + e^{-M}\right)$$

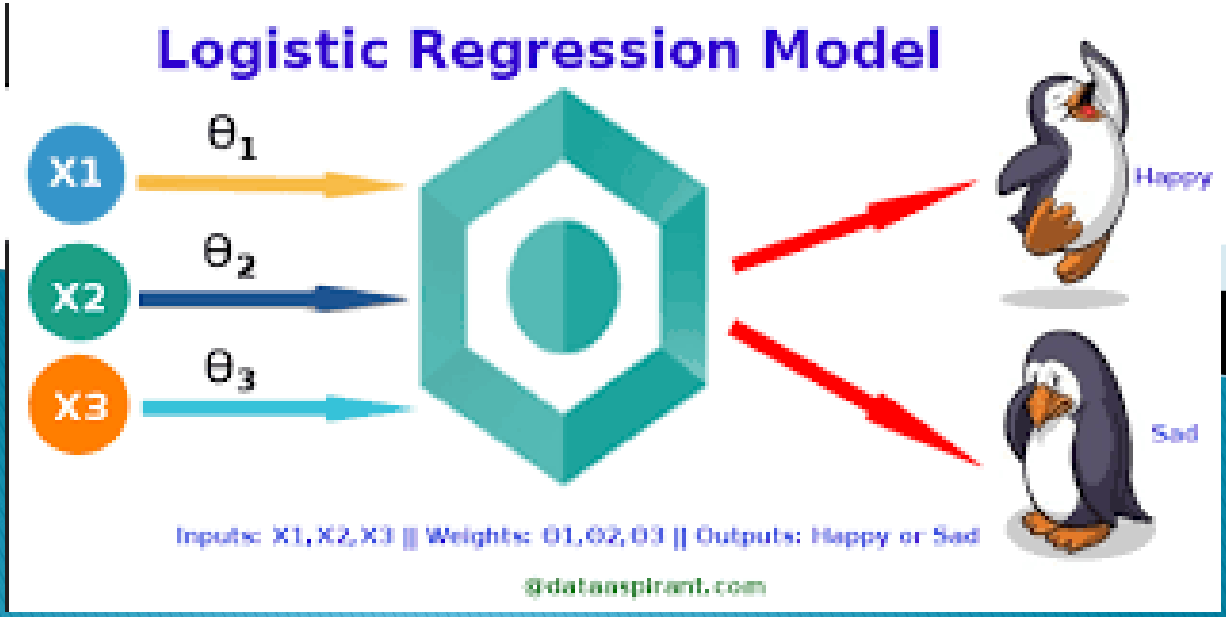
Если мы будем оптимизировать линейную модель классификации соответствующим функционалом, то она будет называться **логистической регрессией**

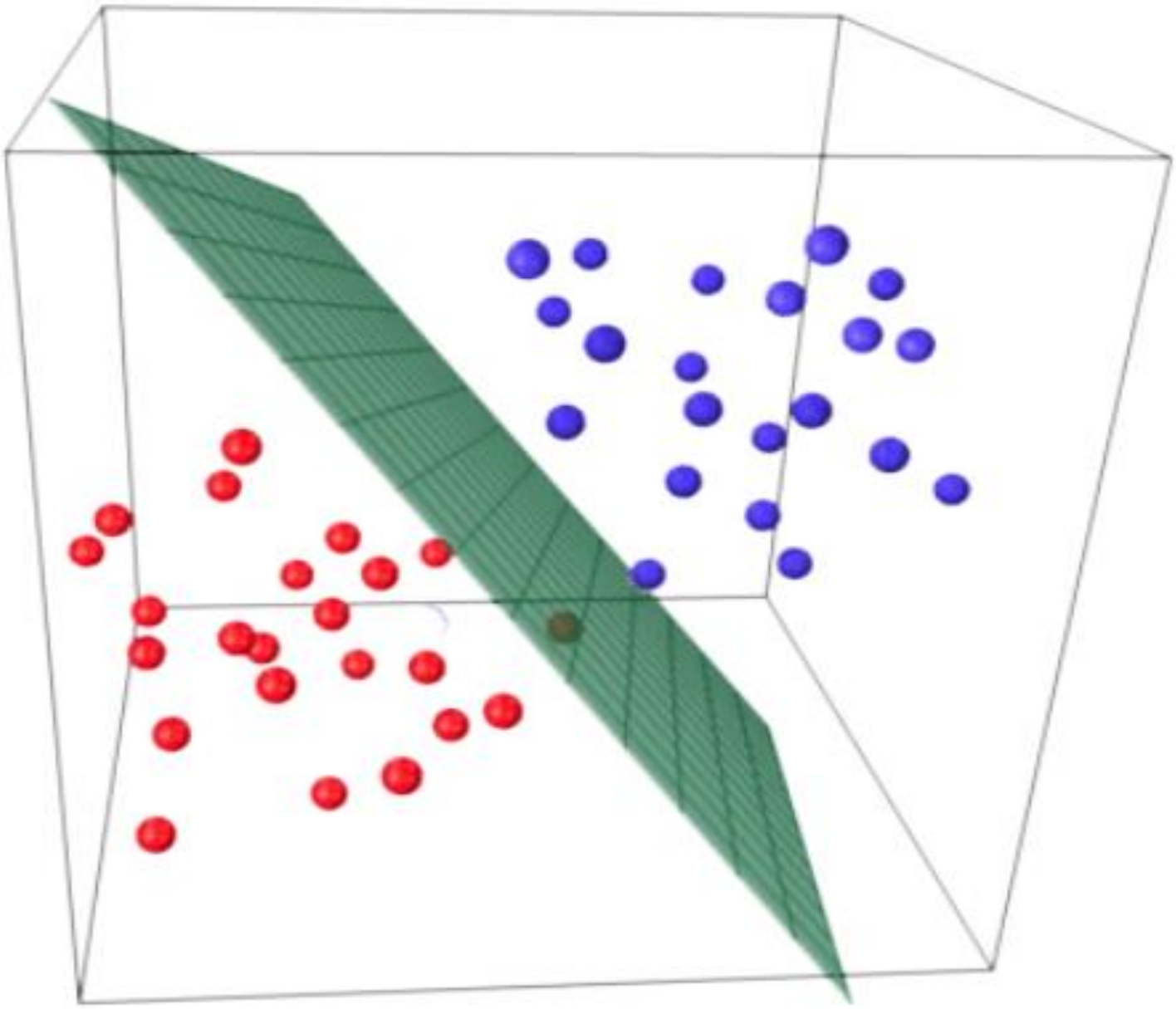
$$\sum_{i=1}^{\ell} \log\left(1 + \exp(-y_i \langle w, x_i \rangle)\right)$$

Логистическая регрессия — это алгоритм классификации машинного обучения, используемый для прогнозирования вероятности категориальной зависимой переменной.
В логистической регрессии зависимая переменная является бинарной переменной, содержащей данные, закодированные как 1 (да, успех и т.п.) или 0 (нет, провал и т.п.).

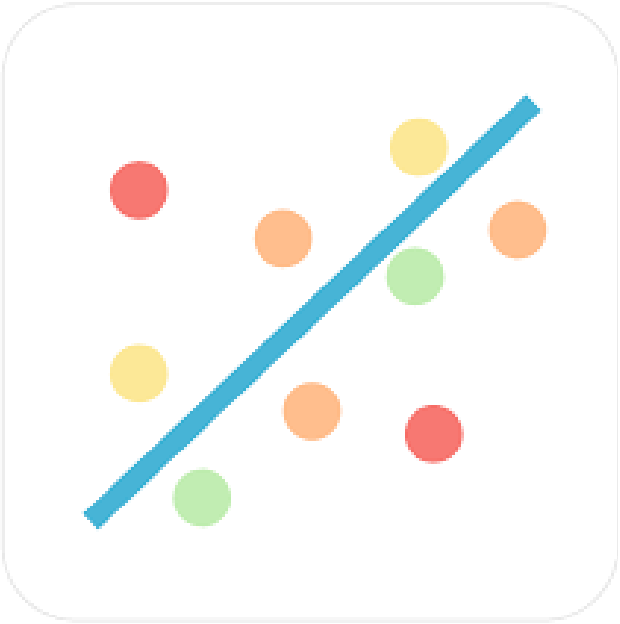
Логистическая регрессия — статистическая модель, используемая для прогнозирования вероятности возникновения некоторого события путём его сравнения с логистической кривой. Эта регрессия выдаёт ответ в виде вероятности бинарного события (1 или 0).

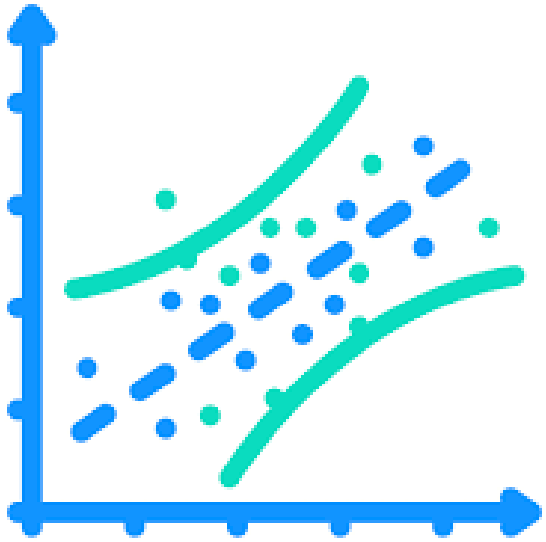
Логистическая регрессия – это разновидность множественной регрессии, предназначенная для классификации записей на основании значений входных полей. При этом выходная переменная является категориальной или бинарной (т.е. может принимать только два значения).

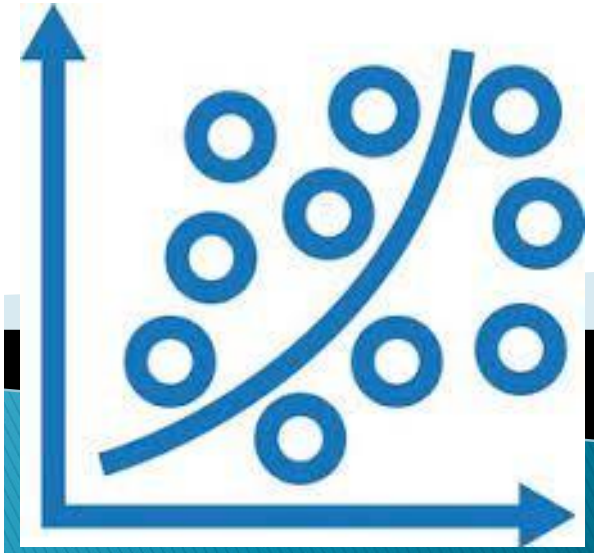




Указанная разделяющая плоскость называется линейным дискриминантом, так как она является линейной с точки зрения своей функции, и позволяет модели производить разделение, дискриминацию точек на различные классы.







Сценарии использования:







Антиспам	Обнаружение мошеннических банковских транзакций	Определение злокачественных опухолей
Признаки: - Характеристики отправителя - Наполнение письма - Содержание в письме стандартных паттернов - Время и место отправления письма Таргет: 0 - Письмо относится к спаму 1 - Письмо не относится к спаму	Признаки: - Характеристики отправителя - Характеристики получателя - Характеристики транзакции: время, гео, устройство Таргет: 0 - Письмо относится к спаму 1 - Письмо не относится к спаму	Признаки: - Характеристики пациента - Характеристики организма в целом - Физические хар-ки органа - Результаты лабораторных исследований Таргет: 0 - Опухоль не злокачественная 1 - Опухоль злокачественная

Условия логистической регрессии:

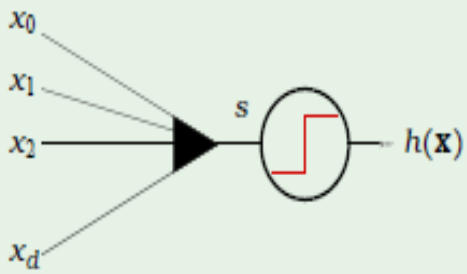
- Бинарная логистическая регрессия требует, чтобы зависимая переменная также была бинарной
- Использоваться должны только значимые переменные.
- Независимые переменные должны быть независимы друг от друга.
- Независимые переменные связаны линейно с логарифмическими коэффициентами.
- Логистическая регрессия требует больших размеров выборки.

Линейные модели

$$s = \sum_{i=0}^{d} w_i x_i$$

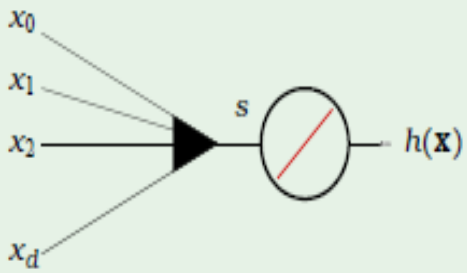
linear classification

$$h(\mathbf{x}) = \text{sign}(s)$$



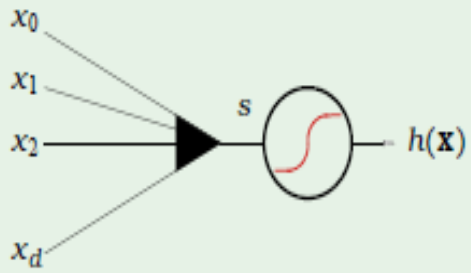
linear regression

$$h(\mathbf{x}) = s$$



logistic regression

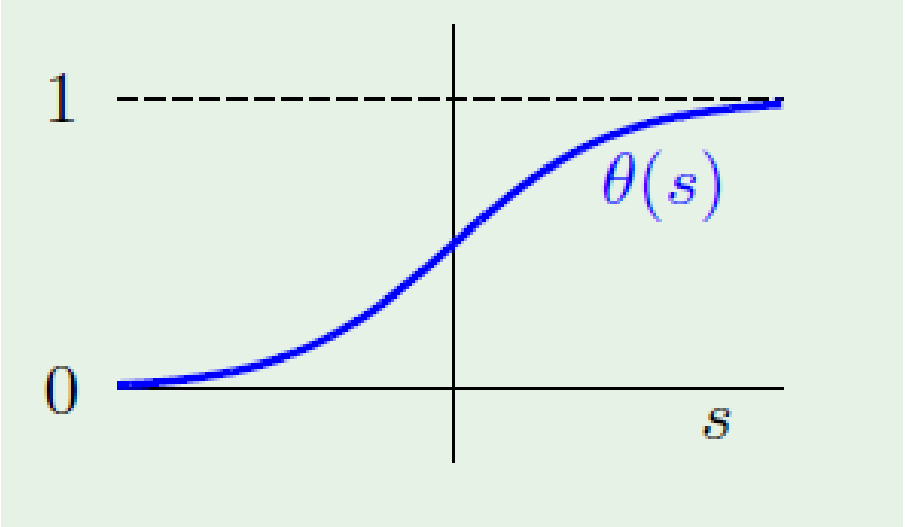
$$h(\mathbf{x}) = \theta(s)$$



Логистическая функция

- «Мягкий порог»

$$\theta(s) = \frac{e^s}{1 + e^s}$$



- Функция θ интерпретируется как вероятность

- Данные (x,y), где y бинарная функция, генерируются «шумной» ЦФ

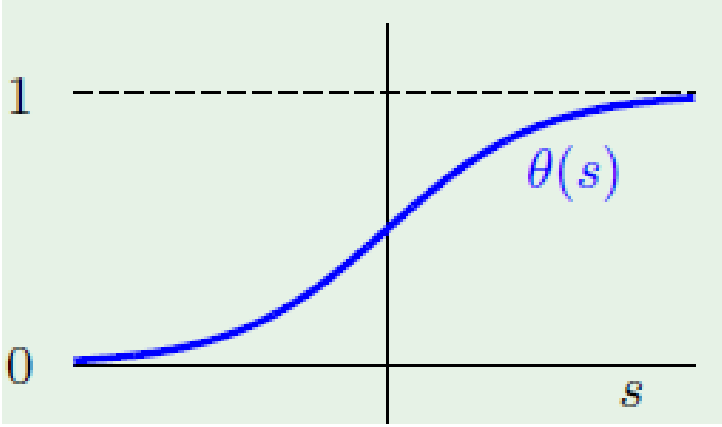
$$P(y \mid \mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{for } y = +1; \\ 1 - f(\mathbf{x}) & \text{for } y = -1. \end{cases}$$

- Функция $f : \mathbb{R}^d \rightarrow [0, 1]$ – вероятность
- Изучаем

$$g(\mathbf{x}) = \theta(\mathbf{w}^\top \mathbf{x}) \approx f(\mathbf{x})$$

Мера ошибки. Метод максимального правдоподобия

$$P(y \mid \mathbf{x}) = \begin{cases} h(\mathbf{x}) & \text{for } y = +1; \\ 1 - h(\mathbf{x}) & \text{for } y = -1. \end{cases}$$



$$P(y \mid \mathbf{x}) = \theta(y\ \mathbf{w}^{\mathsf{T}}\mathbf{x})$$

$$\theta(-s) = 1 - \theta(s)$$

Правдоподобие $\mathcal{D} = (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$ определяется

$$\prod_{n=1}^{N} P(y_n \mid \mathbf{x}_n) = \prod_{n=1}^{N} \theta(y_n \mathbf{w}^{\mathsf{T}}\mathbf{x}_n)$$

Максимизируем правдоподобие

- Минимизируем

$$-\frac{1}{N}\ln\left(\prod_{n=1}^{N}\theta(y_n\,\mathbf{w}^{\mathsf{T}}\,\mathbf{x}_n)\right)$$

$$=\frac{1}{N}\sum_{n=1}^{N}\ln\left(\frac{1}{\theta(y_n\,\mathbf{w}^{\mathsf{T}}\,\mathbf{x}_n)}\right)$$

$$E_{\text{in}}(\mathbf{w})=\frac{1}{N}\sum_{n=1}^{N}\underbrace{\ln\left(1+e^{-y_n\mathbf{w}^{\mathsf{T}}\mathbf{x}_n}\right)}_{\text{e}\left(h(\mathbf{x}_n),y_n\right)}$$

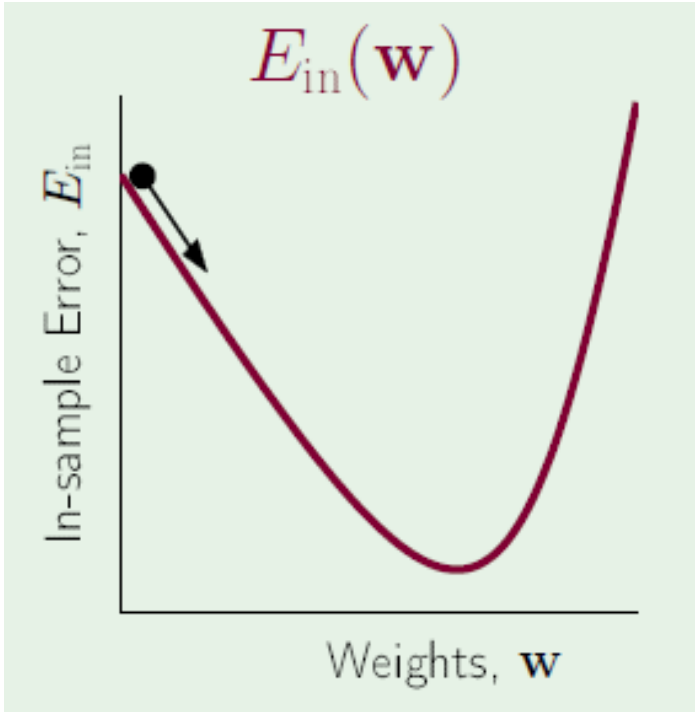
- «кросс-энтропийная» ошибка

Как минимизировать Ein?

$$E_{\text{in}}(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N} \ln\left(1 + e^{-y_n \mathbf{w}^{\mathsf{T}} \mathbf{x}_n}\right)$$

- Метод градиентного спуска

$$\mathbf{w}(1) = \mathbf{w}(0) + \eta\, \hat{\mathbf{v}}$$



1. Медицина

- **Задача:** прогноз наличия заболевания
  Пример: предсказать, есть ли у пациента диабет (да/нет)
- **Признаки (X):**
  возраст
  индекс массы тела (BMI)
  уровень глюкозы
  давление
- **Целевая переменная (y):**
  1 — диабет есть
  0 — диабета нет
- 👉 Используется для:
  ранней диагностики
  оценки риска заболеваний
  медицинского скрининга

2. Банковская сфера (кредитный скоринг)

- **Задача:** оценка вероятности дефолта клиента
- **Признаки:**
  доход
  кредитная история
  возраст
  текущие долги
- **Цель:**
  1 — клиент не вернёт кредит
  0 — клиент надёжный
- 👉 Применение:
  одобрение/отклонение кредита
  расчёт кредитного рейтинга
  управление рисками

3. Маркетинг

- **Задача:** предсказать, купит ли клиент продукт
- **Признаки:**
  возраст
  пол
  история покупок
  клики по рекламе
- **Цель:**
  1 — купит
  0 — не купит
- 👉 Используется для:
  таргетированной рекламы
  сегментации клиентов
  повышения конверсии

4. Кибербезопасность и сети

- **Задача:** обнаружение сетевых атак
- **Примеры:**
  спам / не спам
  нормальный трафик / вредоносный
  успешная аутентификация / взлом
- **Признаки:**
  количество пакетов
  частота запросов
  IP-репутация
  размер трафика
- 👉 Применение:
  IDS/IPS системы
  антиспам
  fraud detection

5. Образование

- **Задача:** прогноз успеваемости студентов
- **Признаки:**
  посещаемость
  баллы за задания
  активность в LMS
  время на платформе
- **Цель:**
  1 — студент сдаст курс
  0 — не сдаст
- 👉 Используется для:
  персонализации обучения
  снижения отсева

6. Производство и IoT

- **Задача:** предсказать отказ оборудования
- **Признаки:**
  температура
  вибрация
  нагрузка
  время работы
- **Цель:**
  1 — отказ скоро будет
  0 — всё нормально
- 👉 Применение:
  прогнозируемое техническое обслуживание
  цифровые двойники
  промышленный IoT

Когда особенно уместна логистическая регрессия

✅ бинарная классификация

Когда:

- Нужно предсказать один из двух классов (0/1, да/нет, успех/неуспех).
- Важна вероятность принадлежности к классу.

- **Почему подходит:**

Модель напрямую оценивает вероятность:

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-z}}, z = w^T x + b$$

Выход — не просто метка, а **вероятность**, что удобно для:

- риск-анализа
- медицины

Когда особенно уместна логистическая регрессия

- ✅ нужна интерпретируемость

  Это **одна из главных причин** выбора логистической регрессии.

- **Почему она хороша:**

  Каждый коэффициент w_i:

  - показывает **направление влияния**
  - показывает **силу влияния**

- **Пример интерпретации:**

  Если $w_{income} = 0.7$ то:

  👉 увеличение дохода на 1 единицу увеличивает шансы события в $e^{0.7} \approx 2.01$ раза.

Когда особенно уместна логистическая регрессия

✅ небольшие и средние данные

Когда:

- мало наблюдений
- данные не «big data»
- важна устойчивость
- нет огромных вычислительных ресурсов

- **Почему логистическая регрессия хороша:**

  Она:

  - имеет **малую дисперсию**
  - быстро обучается
  - не требует GPU
  - редко переобучается при правильной регуляризации
  - хорошо работает уже на сотнях наблюдений

Когда особенно уместна логистическая регрессия

✅ линейная граница решений

Логистическая регрессия — **линейный классификатор**:

$$w^T x + b = 0$$

Это означает:

👉 она лучше всего работает, когда классы **примерно линейно разделимы**.

Подходит, если:

- классы разделяются гиперплоскостью
- зависимости монотонные
- нет сложных нелинейных взаимодействий

Когда особенно уместна логистическая регрессия

✅ базовая модель-бенчмарк

Это **золотой стандарт baseline-модели.**

Почему её почти всегда запускают первой:

- очень быстрая
- стабильная
- понятная
- даёт сильный baseline
- легко диагностировать проблемы данных

Резюме

- Логистическая регрессия — обучение модели так, что на объектах с близкими прогнозами эти прогнозы стремятся к доле положительных объектов

Спасибо за внимание!