

Методы машинного обучения

Лекция 6

Линейная классификация

Эльвира Зиннурова

elvirazinnurova@gmail.com

НИУ ВШЭ, 2019

Модель линейной классификации

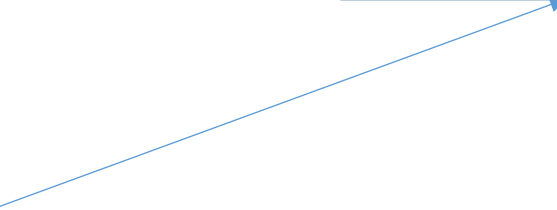
Классификация

- $\mathbb{Y} = \{-1, +1\}$ — бинарная классификация
- -1 — отрицательный класс
- $+1$ — положительный класс
- $a(x)$ должен возвращать одно из двух чисел

Линейная регрессия

$$a(x) = w_0 + \sum_{j=1}^d w_j x^j$$

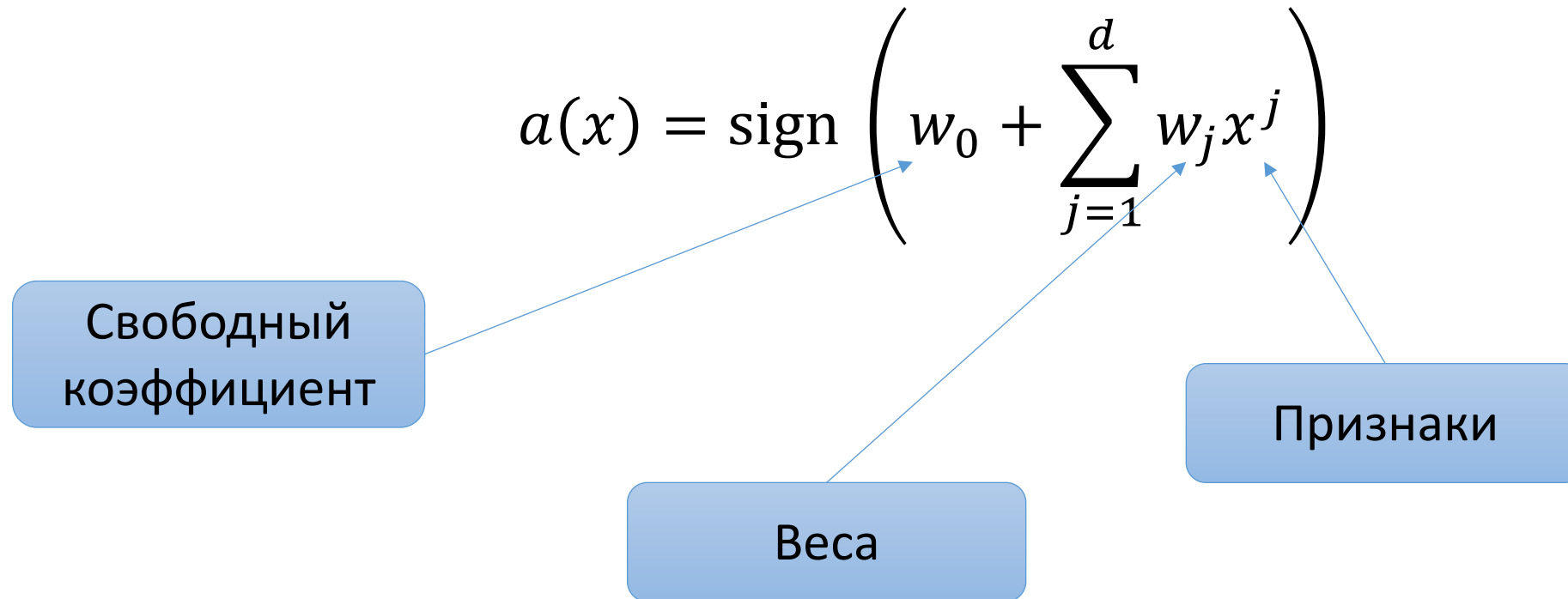
Вещественное
число!



Линейный классификатор

$$a(x) = \text{sign} \left(w_0 + \sum_{j=1}^d w_j x^j \right)$$

Линейный классификатор



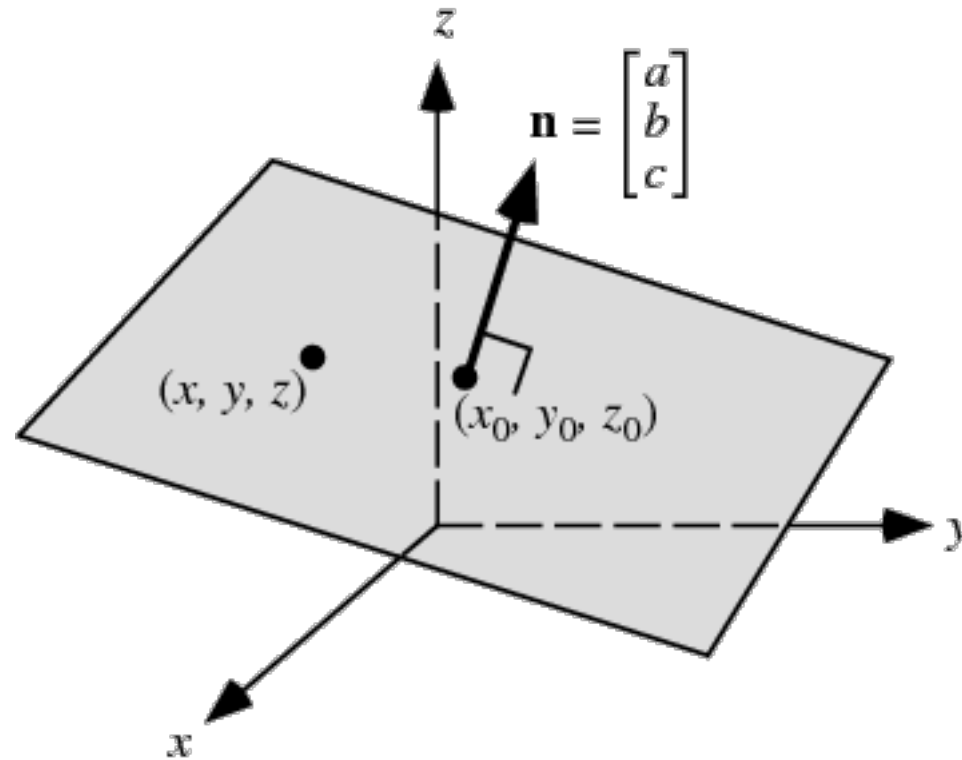
Линейный классификатор

- Добавим единичный признак

$$a(x) = \text{sign} \sum_{j=1}^{d+1} w_j x^j = \text{sign} \langle w, x \rangle$$

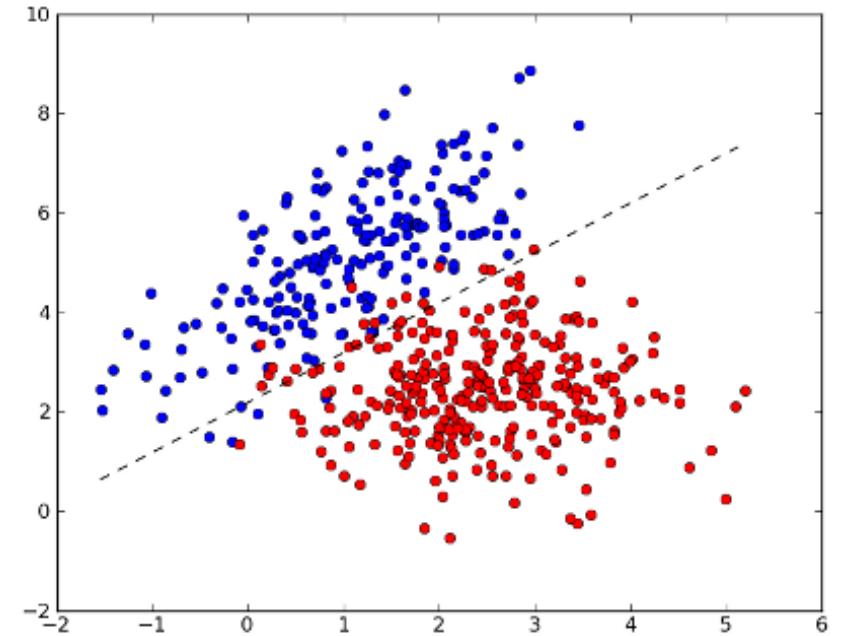
Геометрия линейного классификатора

Уравнение гиперплоскости: $\langle w, x \rangle = 0$



Геометрия линейного классификатора

- Линейный классификатор проводит гиперплоскость
- $\langle w, x \rangle < 0$ — объект «слева» от неё
- $\langle w, x \rangle > 0$ — объект «справа» от неё



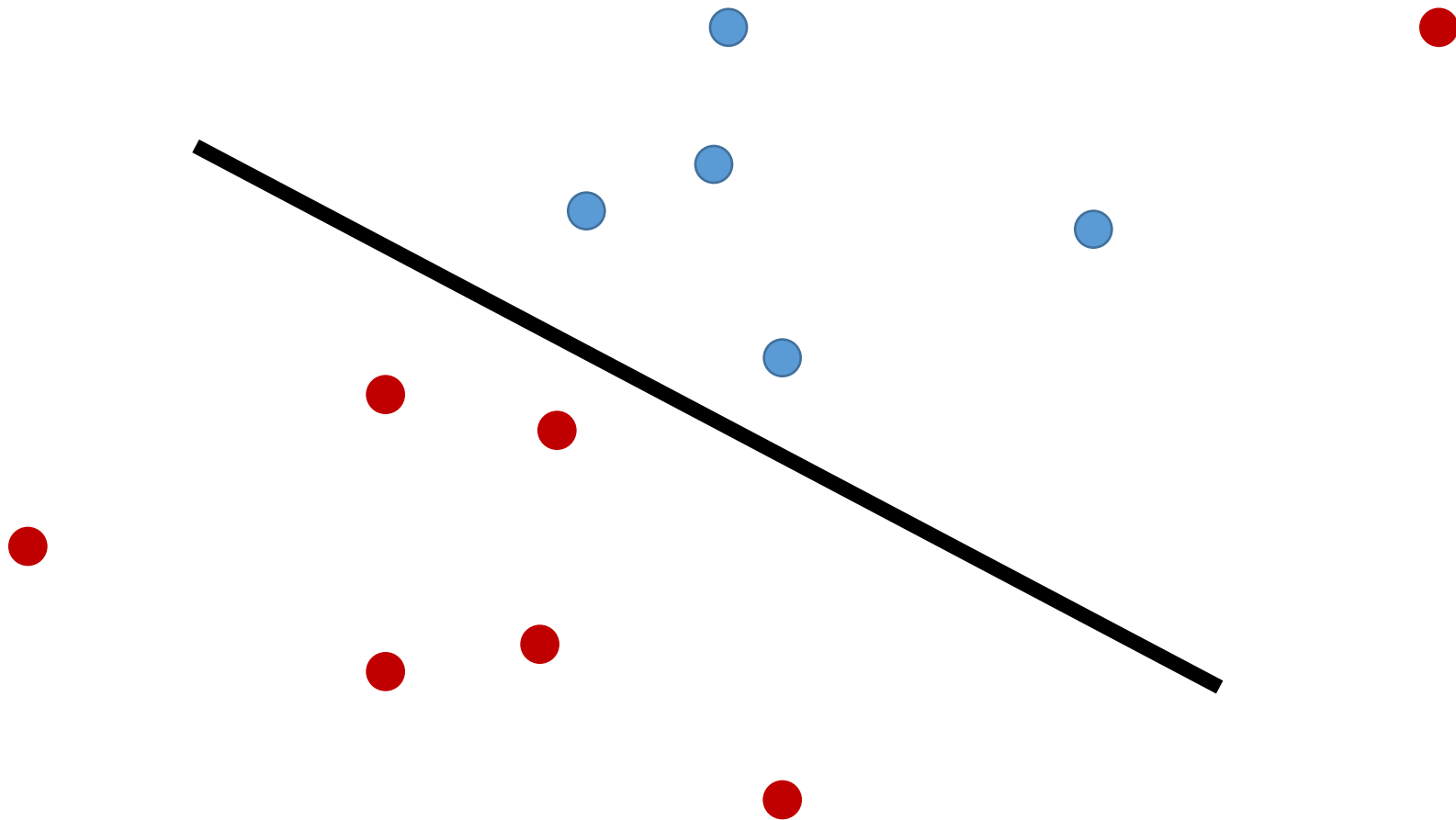
Геометрия линейного классификатора

- Расстояние от точки до гиперплоскости $\langle w, x \rangle = 0$:

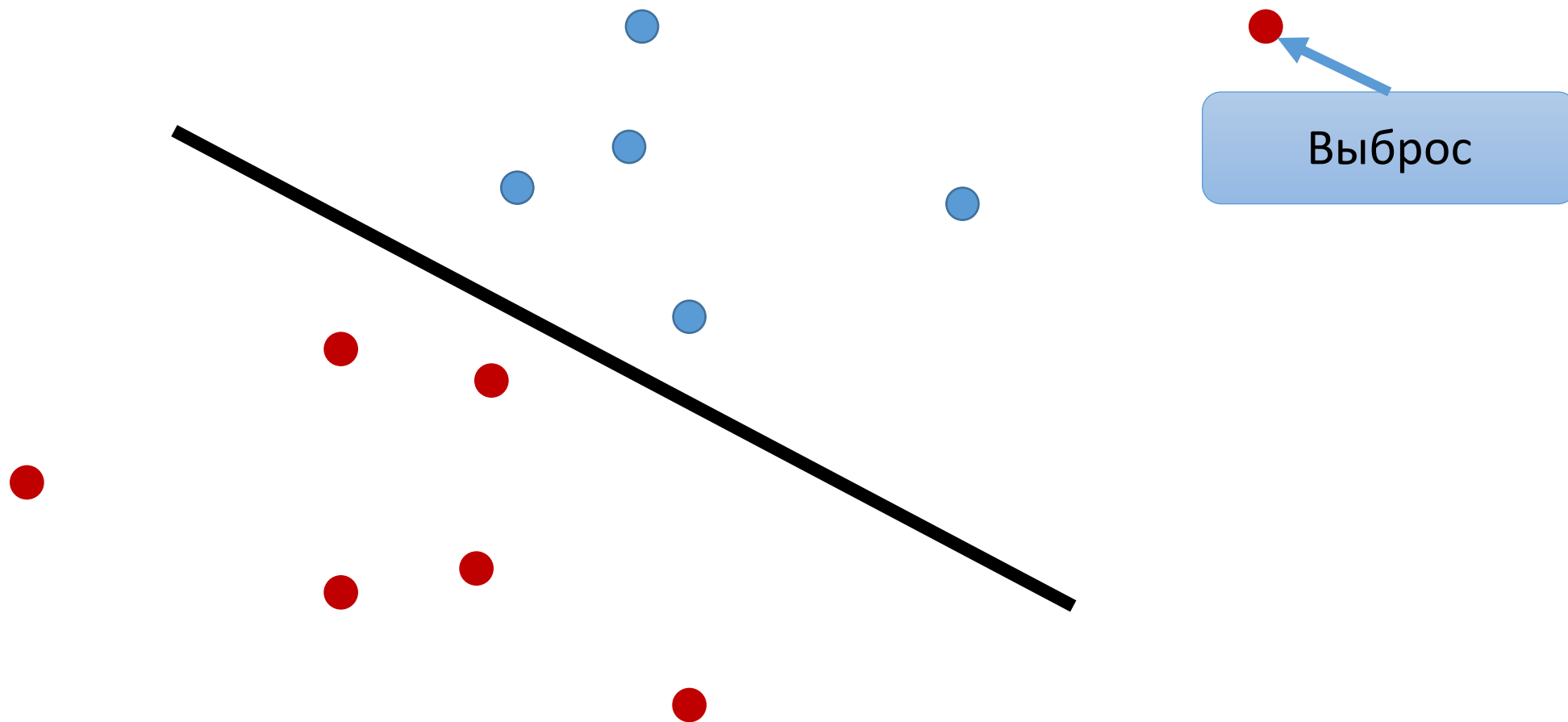
$$\frac{|\langle w, x \rangle|}{\|w\|}$$

- Чем больше $\langle w, x \rangle$, тем дальше объект от разделяющей гиперплоскости

Геометрия линейного классификатора

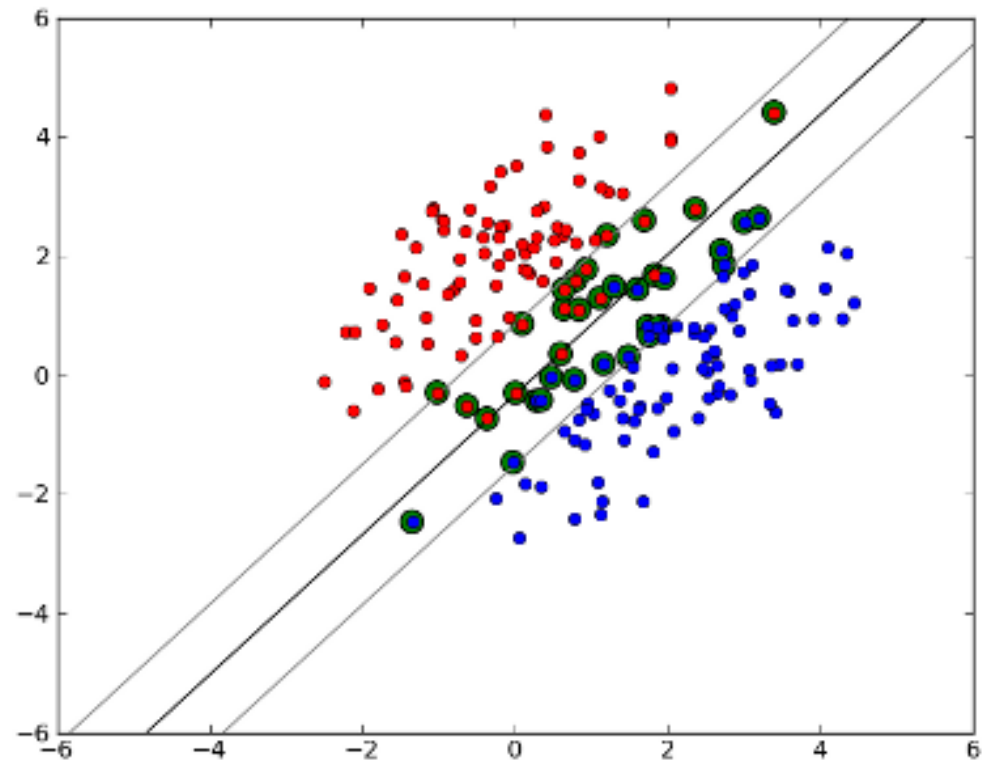


Геометрия линейного классификатора



Отступ

- $M_i = y_i \langle w, x_i \rangle$
- $M_i > 0$ — классификатор дает верный ответ
- $M_i < 0$ — классификатор ошибается
- Чем дальше отступ от нуля, тем больше уверенности



Линейный классификатор

- Линейный классификатор разделяет два класса гиперплоскостью
- Чем больше отступ по модулю, тем дальше объект от гиперплоскости
- Знак отступа говорит о корректности предсказания

Функционал ошибки для
классификации

Линейная регрессия

- Квадратичное отклонение:

$$L(a, y) = (a - y)^2$$

- Абсолютное отклонение:

$$L(a, y) = |a - y|$$

Линейная классификация

- Доля **неправильных** ответов:

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

Линейная классификация

$a(x)$	y
-1	-1
+1	+1
-1	-1
+1	-1
+1	+1

- Доля неправильных ответов:

$$\frac{1}{5} = 0.2$$

Линейная классификация

- Доля **правильных** ответов:

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

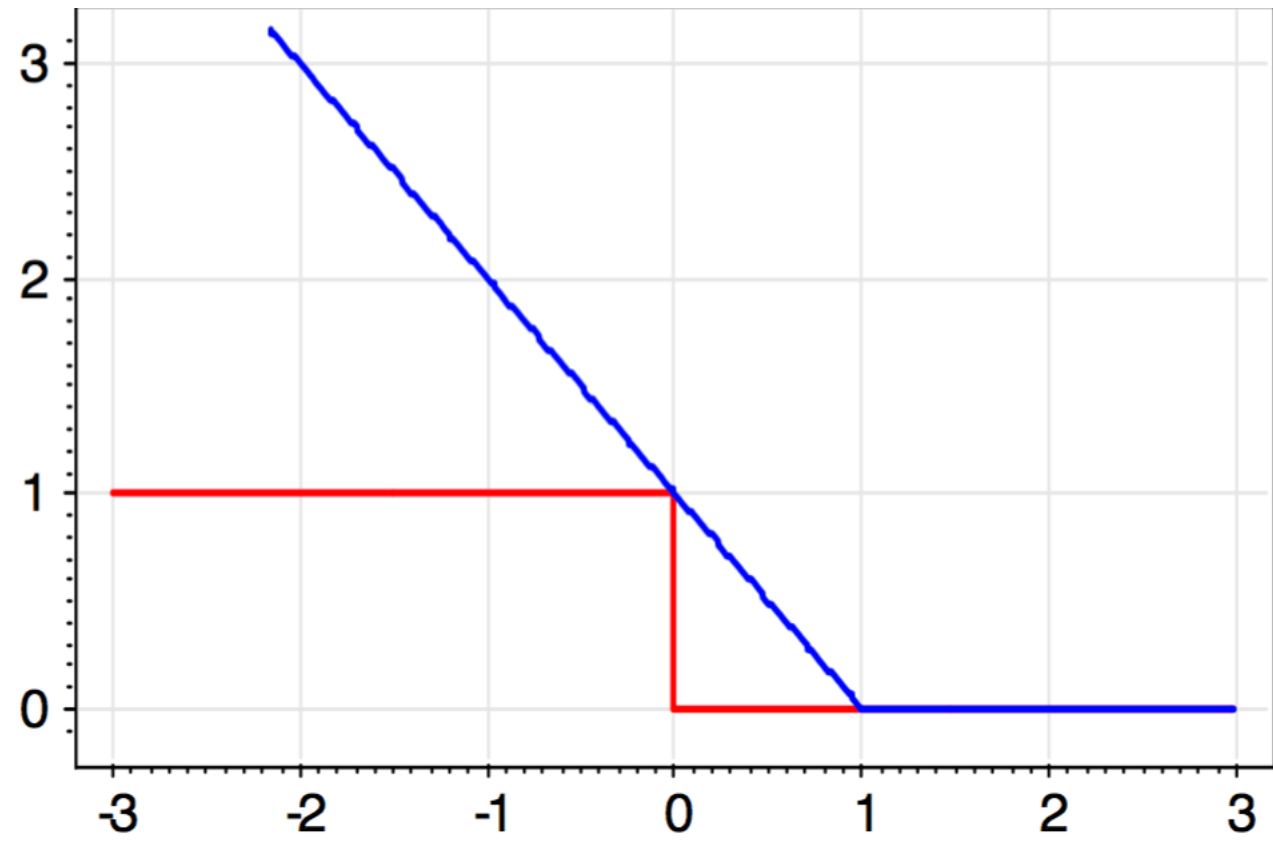
- На английском: **accuracy** (**не** «точность»!)
- Непонятно, как оптимизировать (долю неправильных ответов — тоже)

Линейная классификация

- Доля неправильных ответов (через отступ):

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

Пороговая функция потерь



Линейная классификация

- Доля неправильных ответов:

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

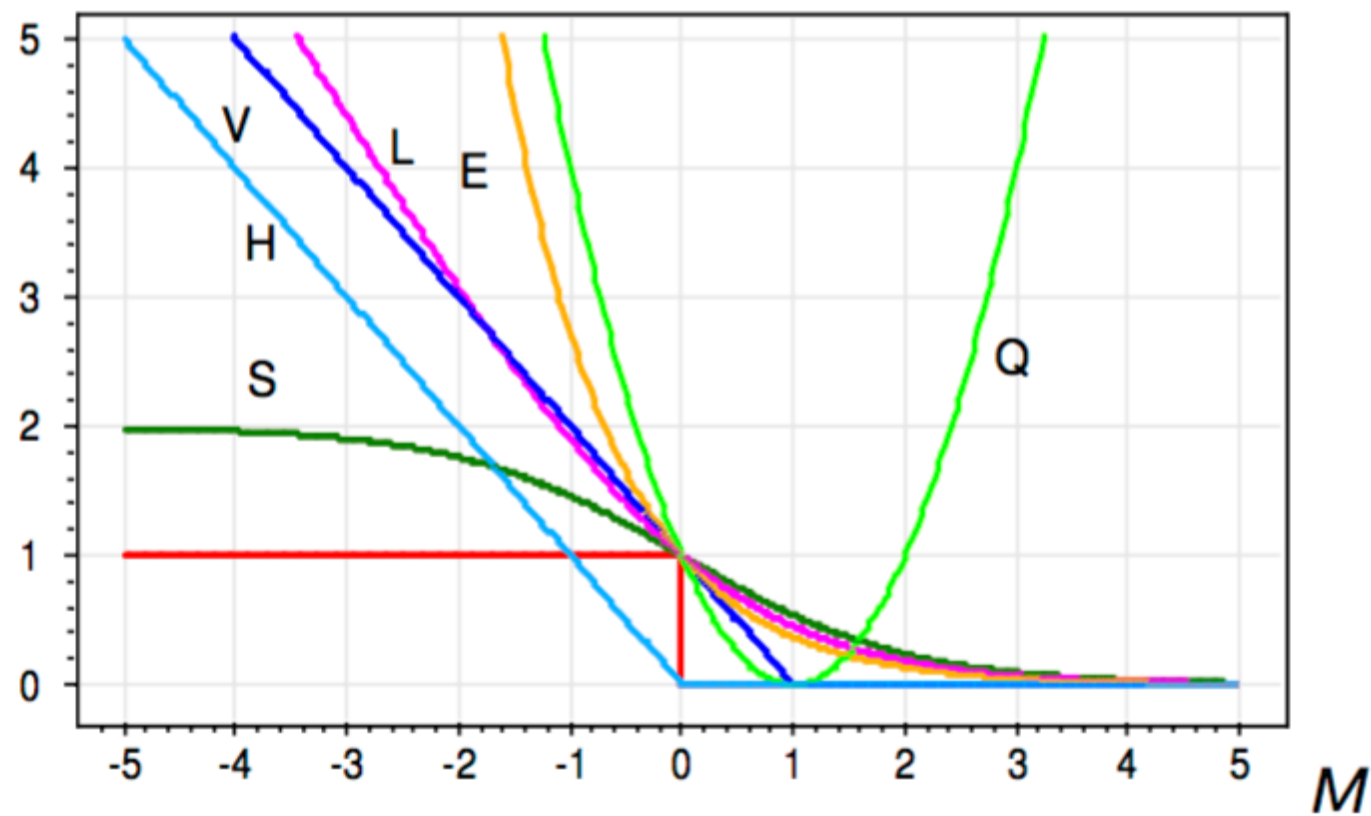
- Разрывная функция
- Все еще непонятно, как оптимизировать

Оценка функции потерь

- Возьмем любую гладкую оценку пороговой функции:

$$[M < 0] \leq \tilde{L}(M) = \tilde{L}(y\langle w, x \rangle)$$

Примеры оценок



Примеры оценок

- $\tilde{L}(M) = \log_2(1 + \exp(-M))$ — логистическая
- $\tilde{L}(M) = \exp(-M)$ — экспоненциальная
- $\tilde{L}(M) = \max(0, 1 - M)$ — кусочно-линейная

Оценка функции потерь

- Возьмем любую гладкую оценку пороговой функции:

$$[M < 0] \leq \tilde{L}(M)$$

- Оценим через нее функционал ошибки:

$$Q(a, X) \leq \tilde{Q}(a, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} \tilde{L}(M_i)$$

Оценка функции потерь

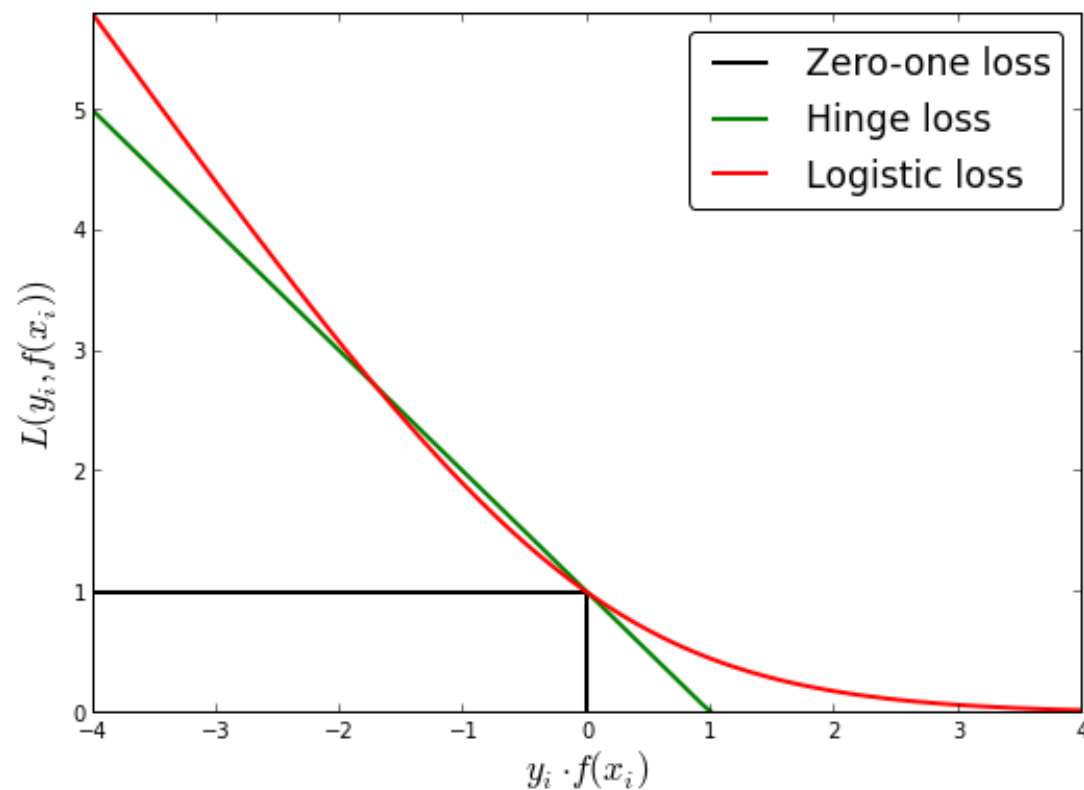
$$\frac{1}{\ell} \sum_{i=1}^{\ell} [M_i < 0] \leq \frac{1}{\ell} \sum_{i=1}^{\ell} \tilde{L}(M_i) \rightarrow \min_a$$

Минимизируем
верхнюю оценку

Надеемся, что доля
ошибок тоже
уменьшится

Примеры оценок

- $\tilde{L}(a, y) = \ln(1 + \exp(-ya))$ — логистическая



Логистическая функция потерь

$$\tilde{Q}(w, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} \log_2(1 + \exp(-y_i \langle w, x_i \rangle))$$

1. Выписали индикатор ошибки через отступ
2. Заменяли пороговую функцию потерь на гладкую функцию

Обучение

- Обучение — с помощью любых методов оптимизации
- Например, градиентный спуск:

$$w^{(t)} = w^{(t-1)} + \eta \frac{1}{\ell} \sum_{i=1}^{\ell} \frac{y_i x_i}{1 + \exp(y_i \langle w, x_i \rangle)}$$

- Борьба с переобучением — регуляризация (так же, как в линейной регрессии), например:

$$\tilde{Q}(w, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} \log_2(1 + \exp(-y_i \langle w, x_i \rangle)) + \lambda \|w\|^2$$

Логистическая регрессия

Логистическая регрессия

$$\frac{1}{\ell} \sum_{i=1}^{\ell} \log_2(1 + \exp(-y_i \langle w, x_i \rangle)) \rightarrow \min_w$$

Оценивание вероятностей

- $P(y = +1 \mid x) = \pi(x)$

Оценивание вероятностей

- Кредитный скоринг
- Стратегия: выдавать кредит только клиентам с $\pi(x) > 0.9$
- 10% невозвращённых кредитов — нормально

Оценивание вероятностей

- Баннерная реклама
- $\pi(x)$ — вероятность, что пользователь кликнет по рекламе
- $c(x)$ — прибыль в случае клика
- $\pi(x)c(x)$ — хотим оптимизировать

Оценивание вероятностей

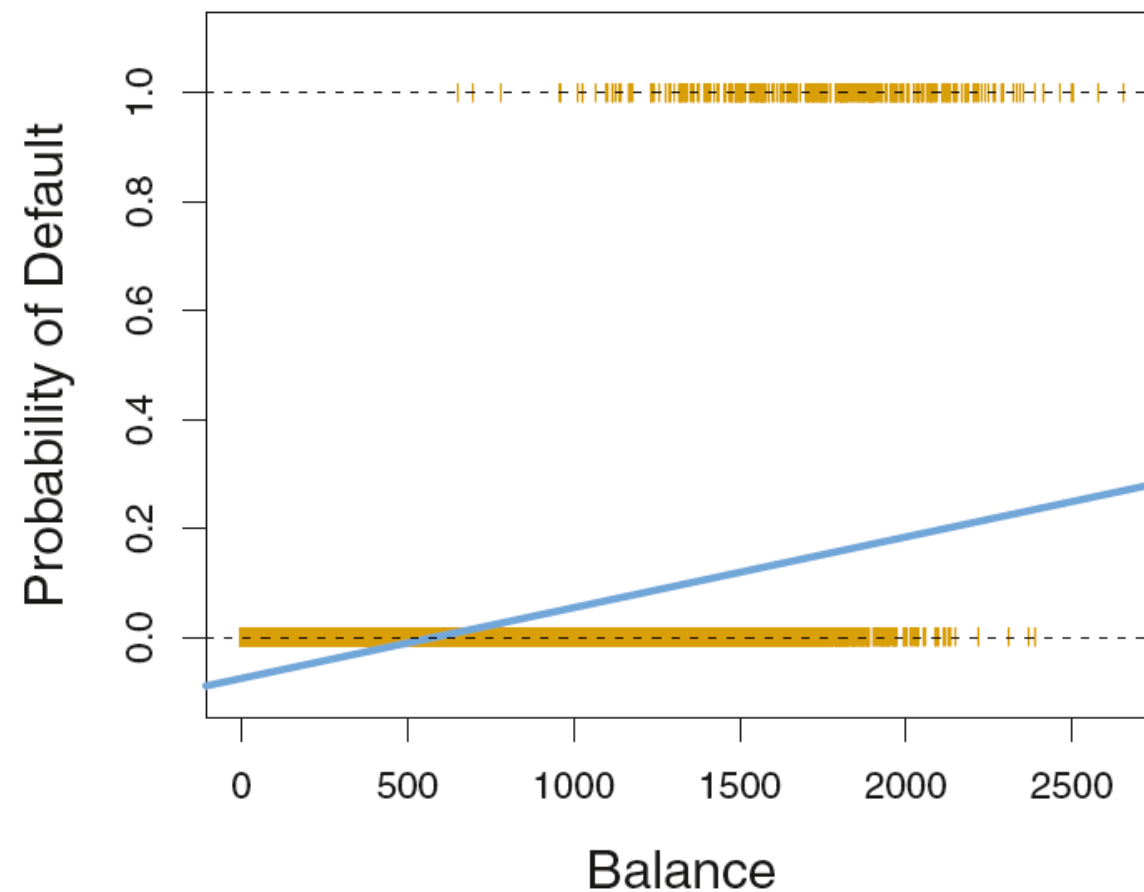
- Прогнозирование оттока клиентов
- Медицинская диагностика
- Поисковое ранжирование (насколько веб-страница соответствует запросу?)

Оценивание вероятностей

- $P(y = +1 \mid x) = \pi(x)$
- $\pi(x)$ — вещественное число
- Классификатор не подходит

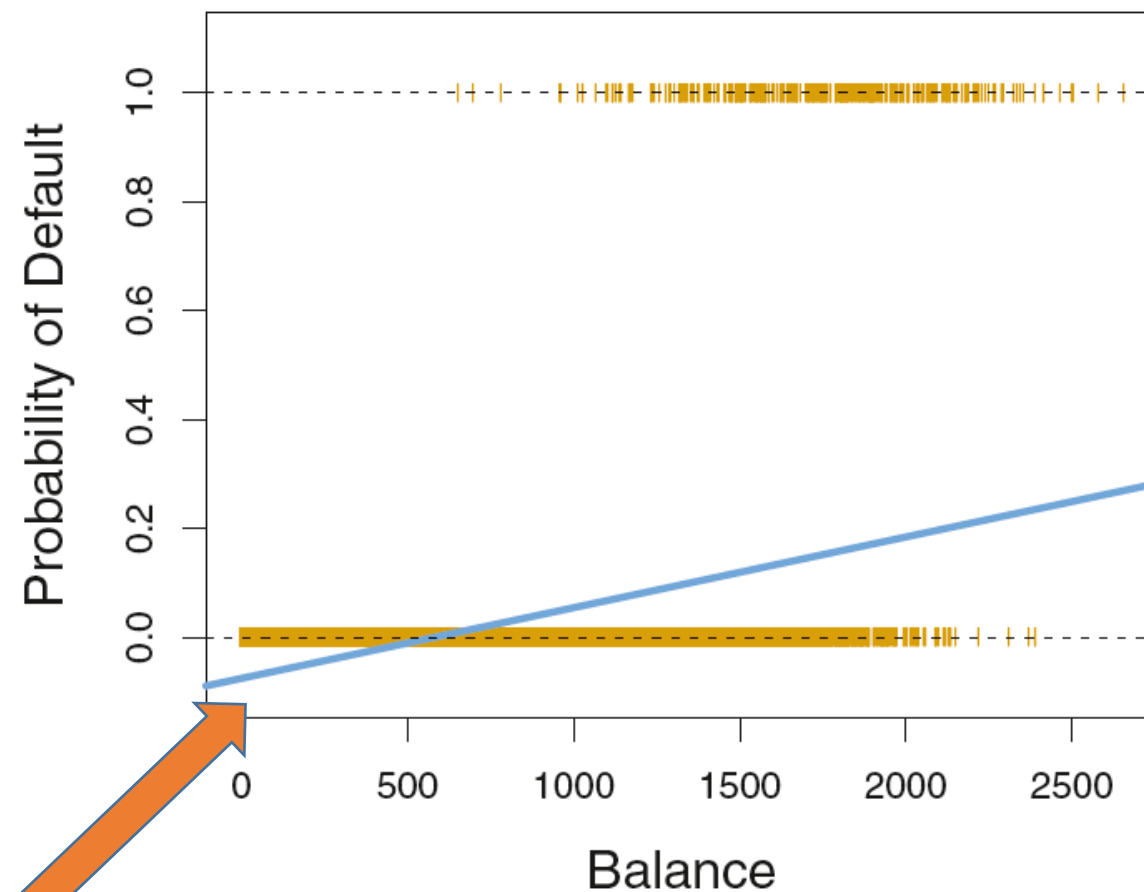
Регрессия?

- $\pi(x) \approx \langle w, x \rangle = w_1 x + w_0$



Регрессия?

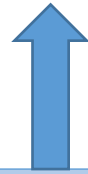
- $\pi(x) \approx \langle w, x \rangle = w_1 x + w_0$



Отрицательная вероятность o_0

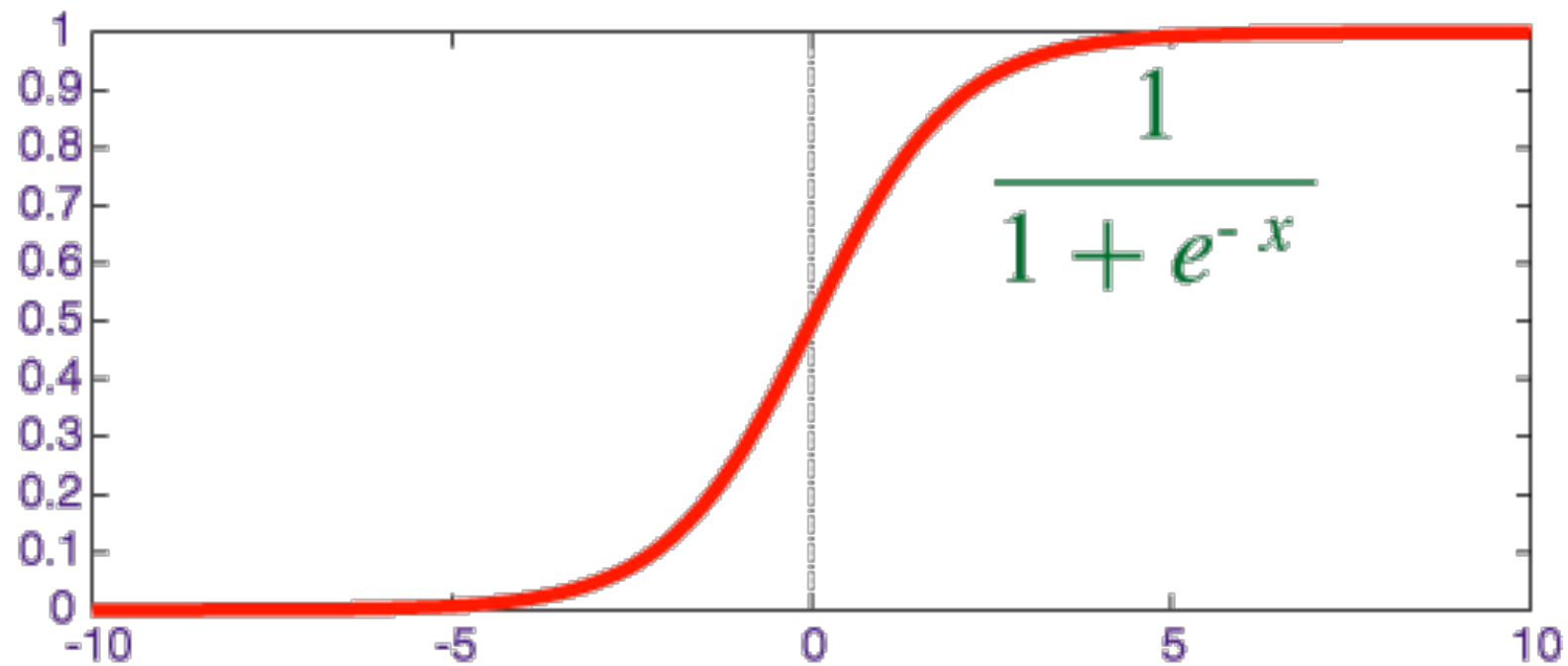
Регрессия?

$$\pi(x) \approx \sigma(\langle w, x \rangle) = \frac{1}{1 + \exp(-\langle w, x \rangle)}$$



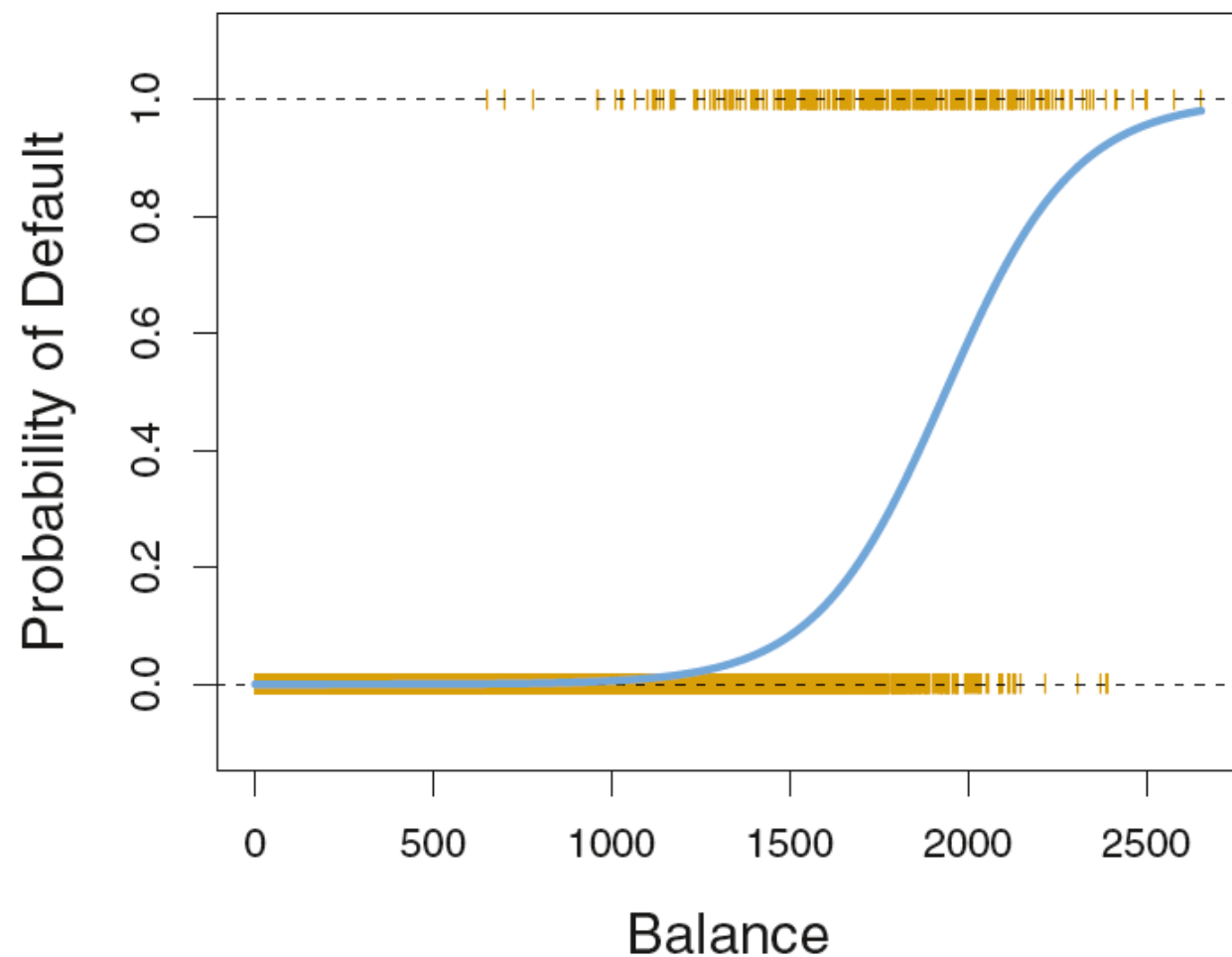
Сигмоида

Сигмоида



Логистическая регрессия

- $\pi(x) \approx \sigma(\langle w, x \rangle)$



Логистическая регрессия

- Как оптимизировать?
- Если $y_i = +1$, то $\langle w, x_i \rangle \rightarrow +\infty$
- Если $y_i = -1$, то $\langle w, x_i \rangle \rightarrow -\infty$

Логистическая регрессия

- Как оптимизировать?
- Если $y_i = +1$, то $\sigma(\langle w, x_i \rangle) \rightarrow 1$
- Если $y_i = -1$, то $\sigma(\langle w, x_i \rangle) \rightarrow 0$

Логистическая регрессия

- Как оптимизировать?
- Если $y_i = +1$, то $\sigma(\langle w, x_i \rangle) \rightarrow 1$
- Если $y_i = -1$, то $\sigma(\langle w, x_i \rangle) \rightarrow 0$

$$\sum_{i=1}^{\ell} \{ [y_i = +1] \sigma(\langle w, x_i \rangle) + [y_i = -1] (1 - \sigma(\langle w, x_i \rangle)) \} \rightarrow \max_w$$

Логистическая регрессия

- Как оптимизировать?
- Если $y_i = +1$, то $\sigma(\langle w, x_i \rangle) \rightarrow 1$
- Если $y_i = -1$, то $\sigma(\langle w, x_i \rangle) \rightarrow 0$

$$\sum_{i=1}^{\ell} \{ [y_i = +1] \sigma(\langle w, x_i \rangle) + [y_i = -1] (1 - \sigma(\langle w, x_i \rangle)) \} \rightarrow \max_w$$

- Слишком слабый штраф
- Если $y_i = +1$ и $\sigma(\langle w, x_i \rangle) = 0$, то штраф = 1

Логистическая регрессия

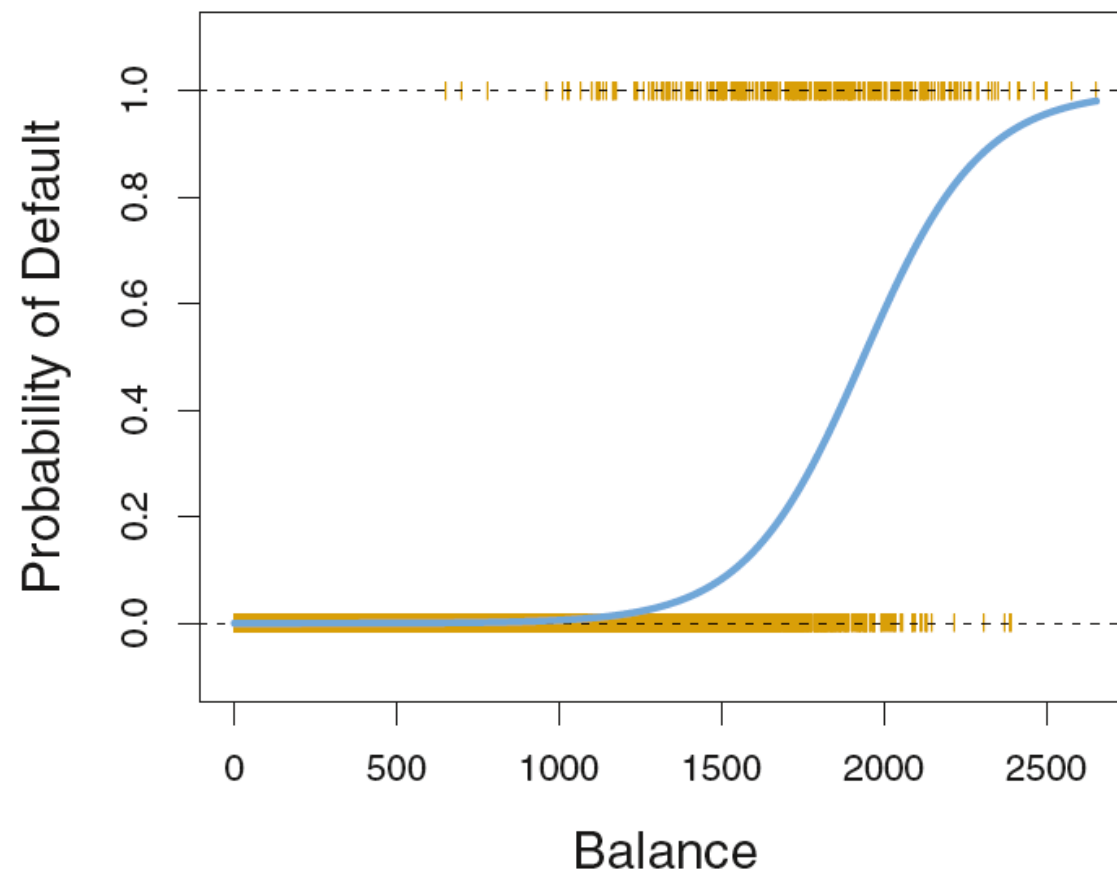
- Как оптимизировать?
- Если $y_i = +1$, то $\sigma(\langle w, x_i \rangle) \rightarrow 1$
- Если $y_i = -1$, то $\sigma(\langle w, x_i \rangle) \rightarrow 0$

$$\sum_{i=1}^{\ell} \{ [y_i = +1] \log_2 \sigma(\langle w, x_i \rangle) + [y_i = -1] \log_2 (1 - \sigma(\langle w, x_i \rangle)) \} \rightarrow \max_w$$

- Если $y_i = +1$ и $\sigma(\langle w, x_i \rangle) = 0$, то штраф $= -\infty$



Логистическая регрессия



Логистическая регрессия

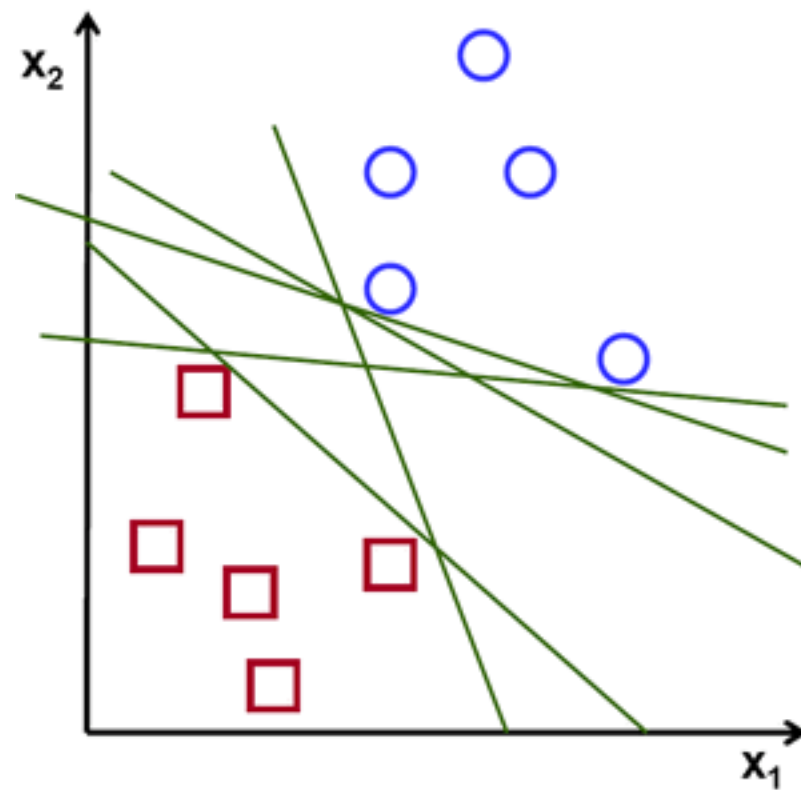
- Если вспомнить арифметику, то получим эквивалентную задачу:

$$\frac{1}{\ell} \sum_{i=1}^{\ell} \log_2(1 + \exp(-y_i \langle w, x_i \rangle)) \rightarrow \min_w$$

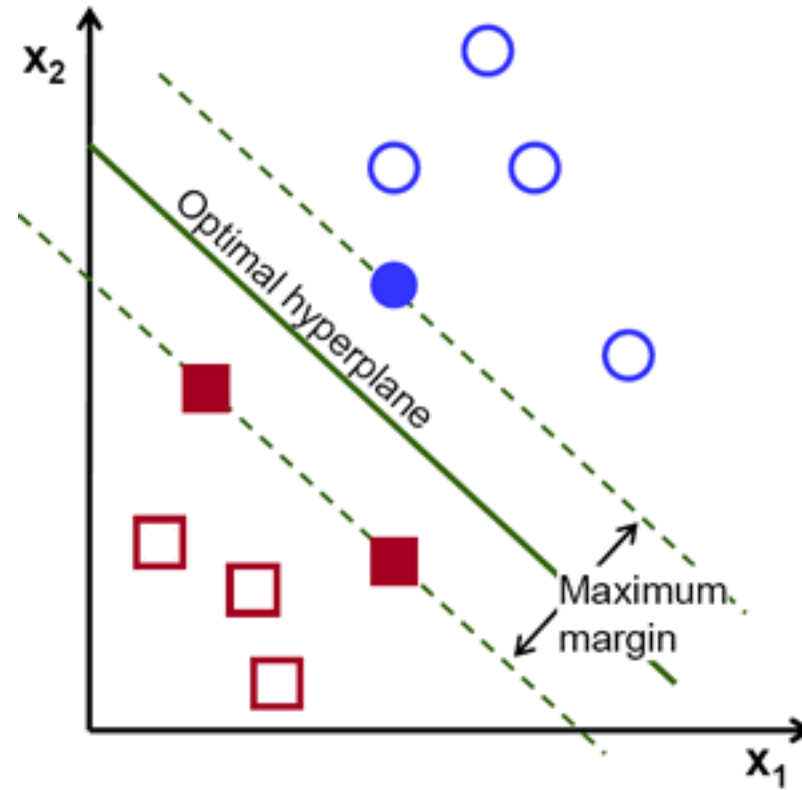
Логистическая регрессия

- Линейная модель классификации: $a(x) = \text{sign } \langle w, x \rangle$
- Оптимизируем логистическую верхнюю оценку для пороговой функции потерь
- Обучение: градиентный спуск
- Позволяет оценивать вероятности: $\pi(x) = \sigma(\langle w, x \rangle)$

SVM



SVM



Резюме

- Линейные классификаторы разделяют классы гиперплоскостью
- Логистическая регрессия — классификация и оценка вероятности