Линейная классификация

Лекция 3

План лекции

- Линейные модели классификации
- Метрики качества классификации
- Логистическая регрессия
- Обучение логистической регрессии
- Метод опорных векторов
- Многоклассовая классификация

Линейные модели классификации

- Обозначим:
  - $\mathbb{X} = \mathbb{R}^d$ – пространство объектов
  - $Y = \{-1, +1\}$ – множество допустимых ответов
  - $X = \{(\vec{x}_i, y_i)\}_{i=1}^{l}$ – обучающая выборка

- Линейная модель классификации:

$$a(\vec{x}) = sign(\langle \vec{w}, \vec{x} \rangle + w_0) = sign\left(\sum_{j=1}^{d} w_j x_j + w_0\right),$$

где $\vec{w} \in \mathbb{R}^d$ – вектор весов, $w_0 \in \mathbb{R}$ – сдвиг, $sign$ – функция знака:

$$sign\, u = \begin{cases} +1, u > 0 \\ 0, u = 0 \\ -1, u < 0 \end{cases}$$

- Если $x_0 = 1$, тогда $a(\vec{x}) = sign\langle \vec{w}, \vec{x} \rangle$

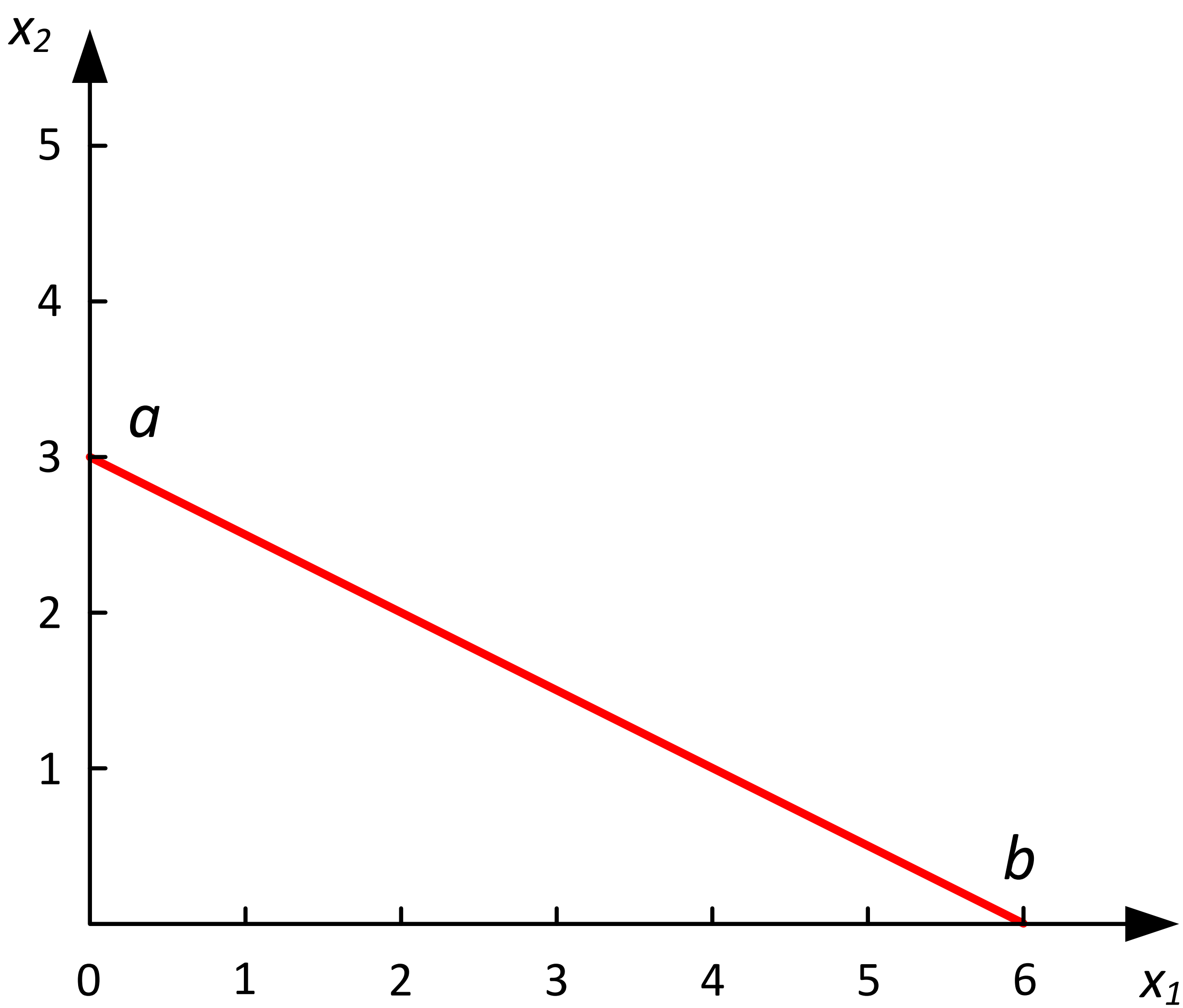
Геометрическая интерпретация

- Линейный классификатор соответствует гиперплоскости с вектором нормали $\vec{w}$
- Величина скалярного произведения $\langle \vec{w}, \vec{x} \rangle$ пропорциональна расстоянию от гиперплоскости до точки $\vec{x}$, а его знак показывает, с какой стороны от гиперплоскости находится данная точка
- Расстояние от точки до гиперплоскости:

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

- Линейный классификатор разделяет пространство на две части с помощью гиперплоскости, и при этом одно полупространство относится к положительному классу, а другое – к отрицательному

Геометрическая интерпретация



Геометрическая интерпретация

- Уравнение прямой по двум точкам:

$$(x_{1a}, x_{2a}) = (0, 3), \qquad (x_{1b}, x_{2b}) = (6, 0)$$

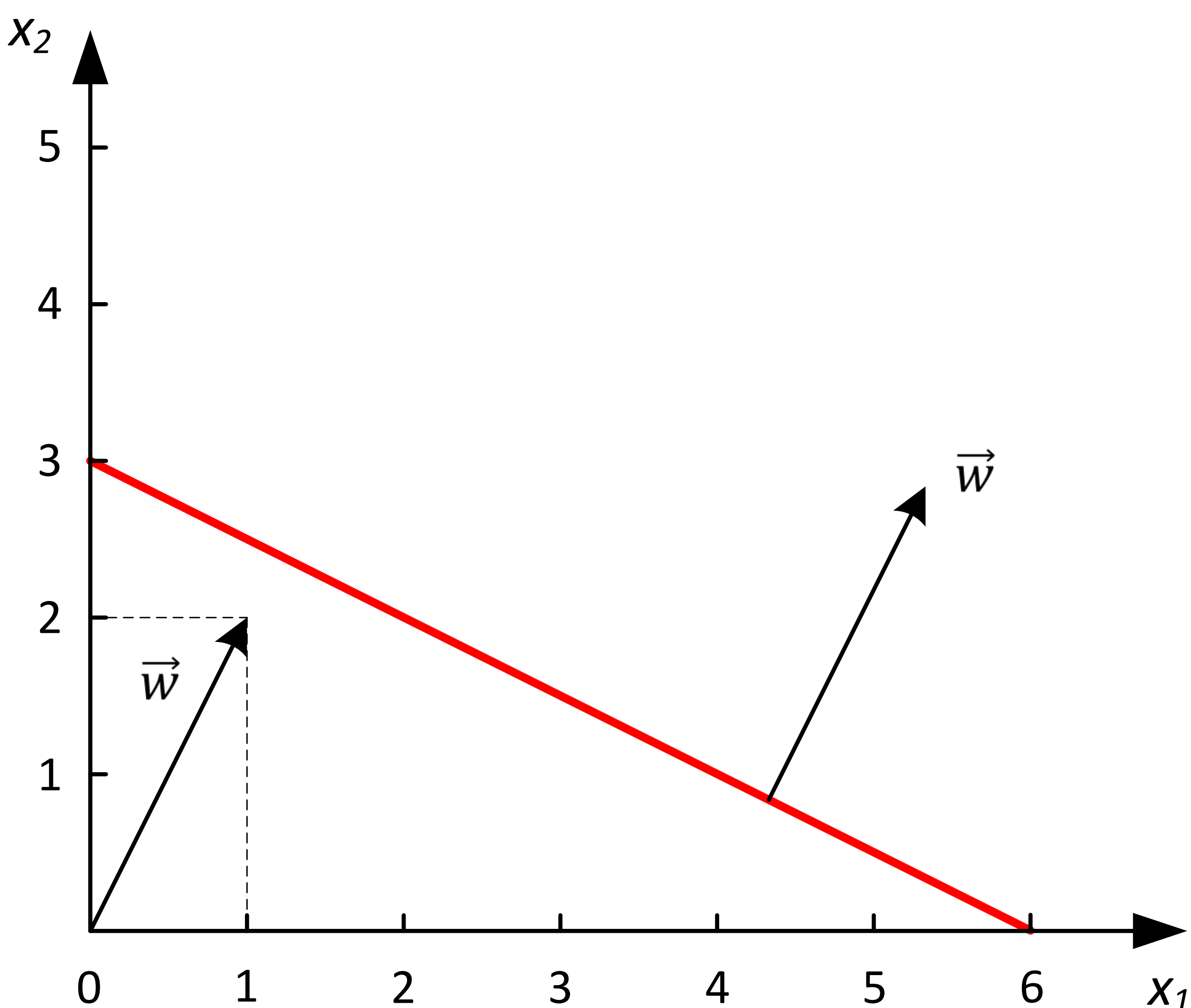
$$(x_{2a} - x_{2b})x_1 + (x_{1b} - x_{1a})x_2 + (x_{1a}x_{2b} - x_{1b}x_{2a}) = 0$$

$$3x_1 + 6x_2 - 18 = 0$$

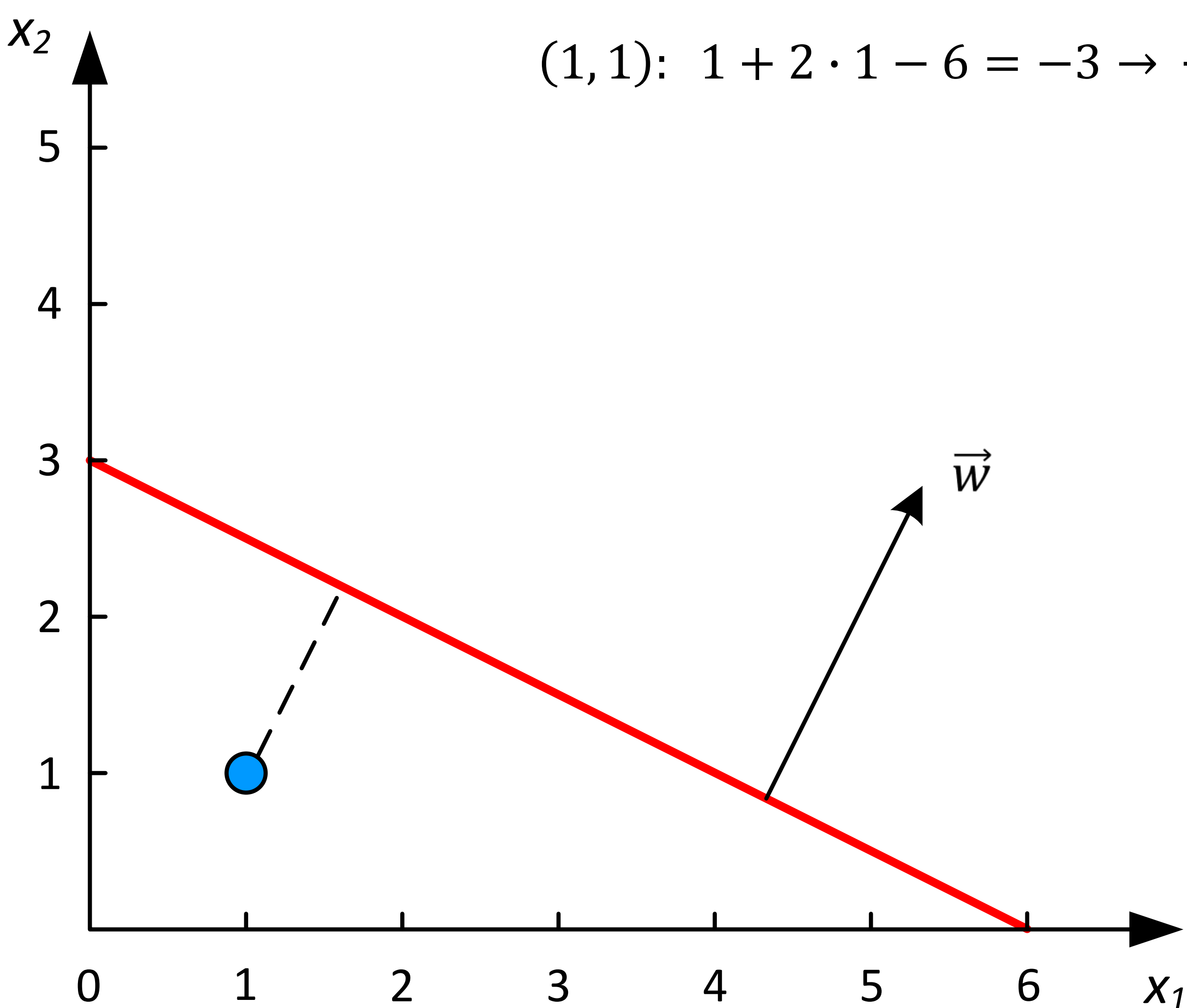
$$x_1 + 2x_2 - 6 = 0$$

$$\vec{w} = (1, 2), \qquad \|\vec{w}\| = \sqrt{1^2 + 2^2} = \sqrt{5} \cong 2.236$$

Геометрическая интерпретация

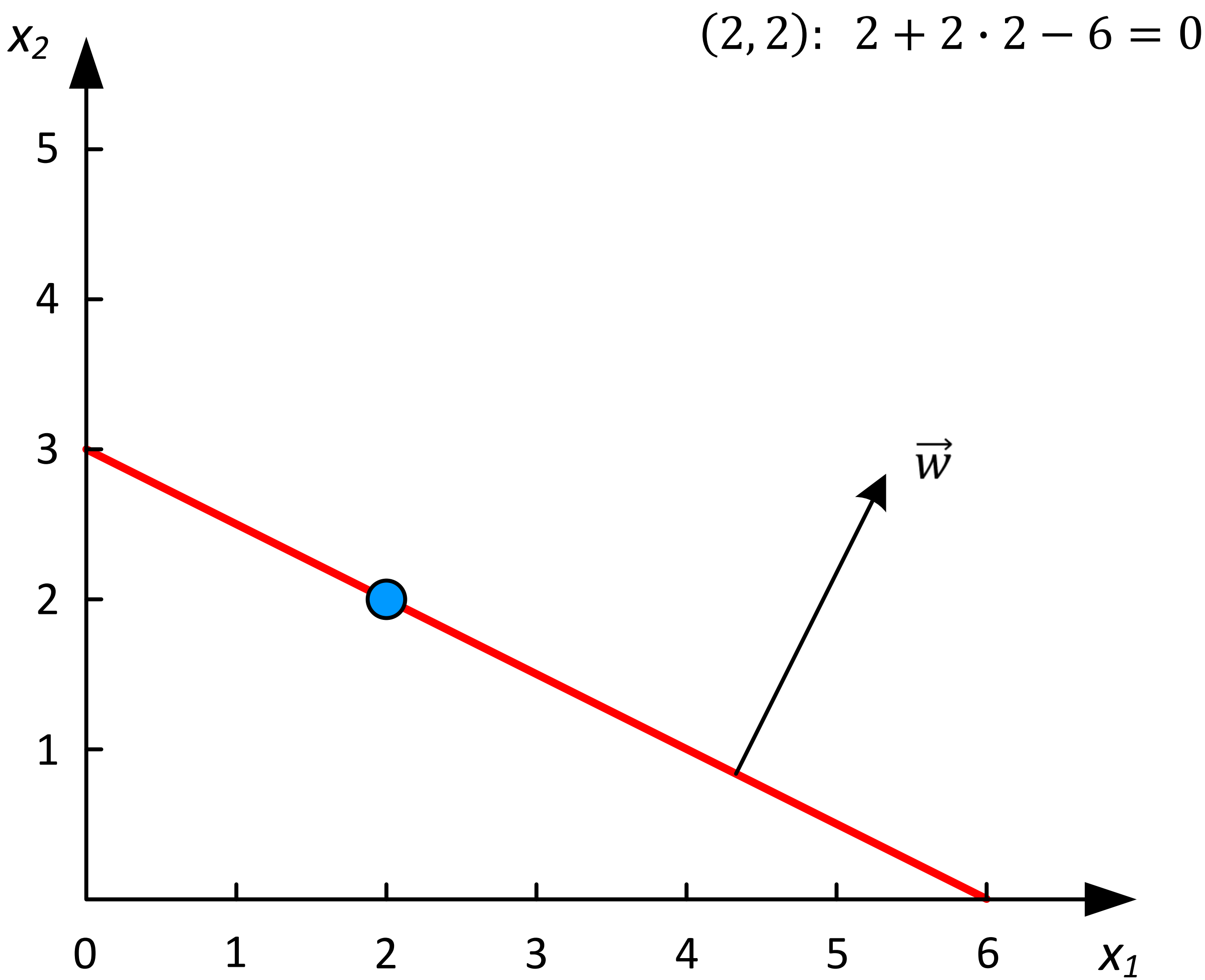


Геометрическая интерпретация



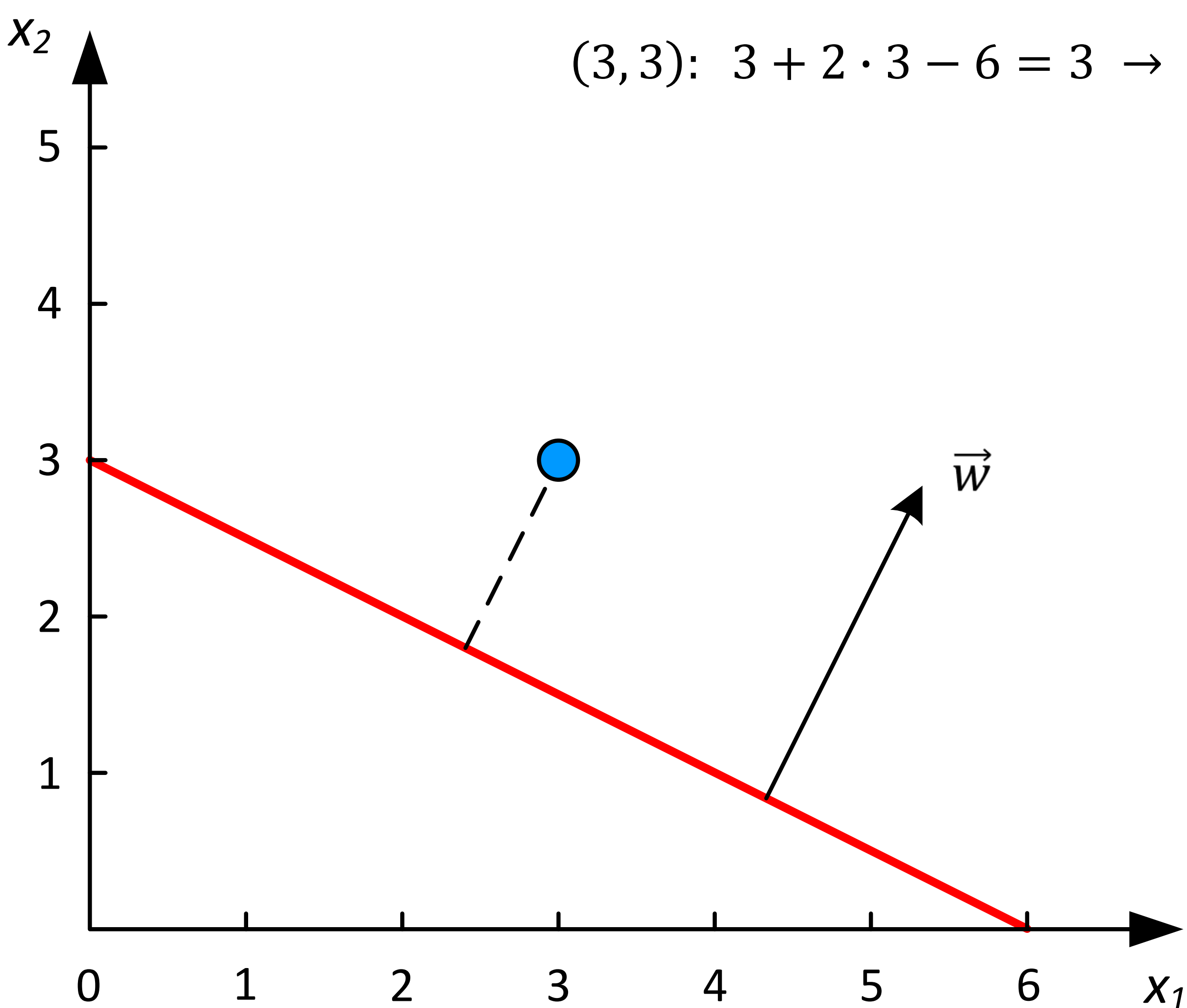
$$(1,1):\ \ 1 + 2 \cdot 1 - 6 = -3 \to \frac{|-3|}{\sqrt{5}} \cong 1.34$$

Геометрическая интерпретация



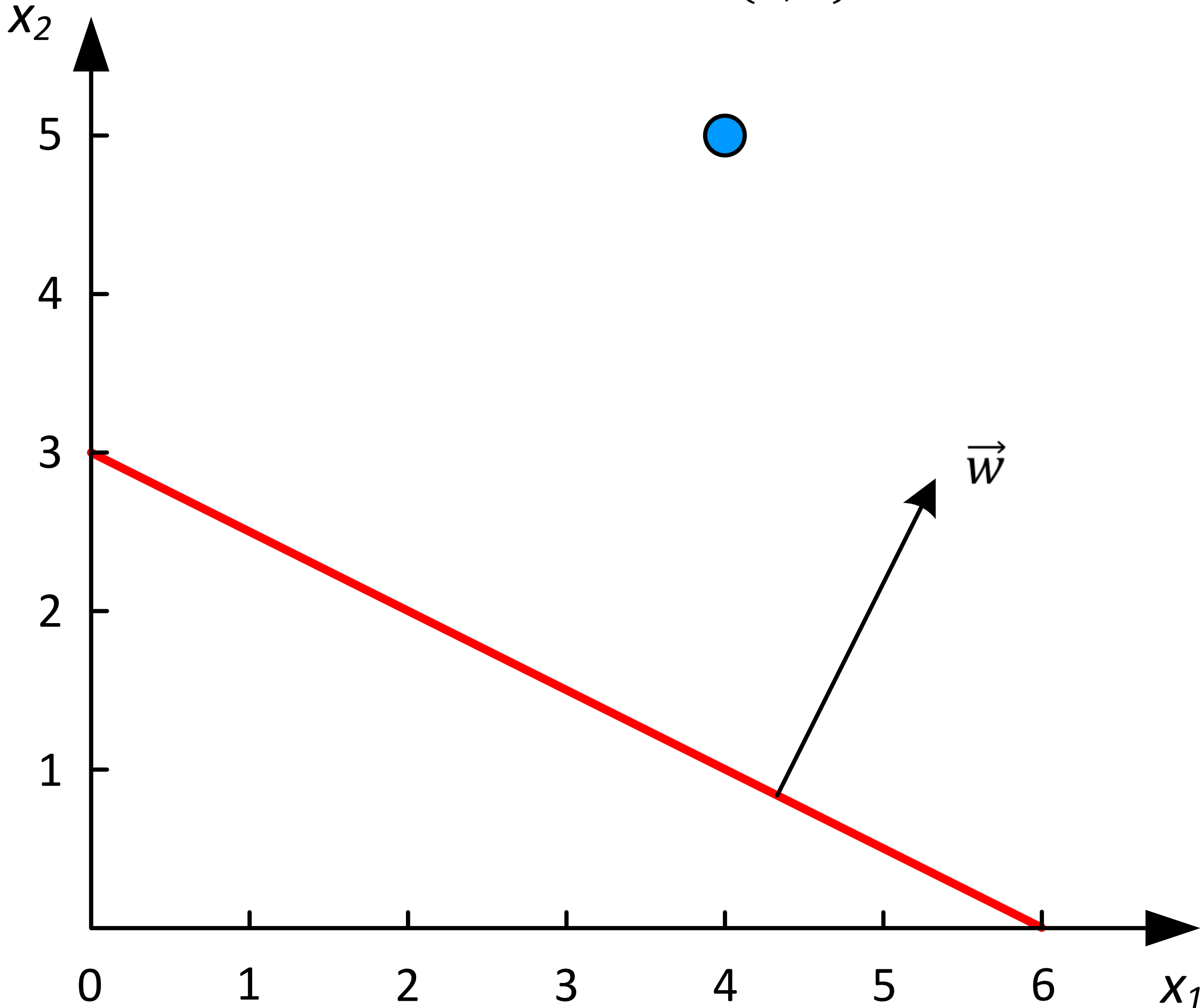
Геометрическая интерпретация

$$(3,3):\ \ 3 + 2 \cdot 3 - 6 = 3\ \rightarrow \frac{|3|}{\sqrt{5}} \cong 1.34$$



Геометрическая интерпретация

$$(4,5):\ \ 4 + 2 \cdot 5 - 6 = 8 \ \rightarrow\ \frac{8}{\sqrt{5}} \cong 3.58$$



Функционалы качества/ошибки

- Доля правильных ответов или *правильность* (accuracy):

$$Q(a, X) = \frac{1}{l}\sum_{i=1}^{l}[a(\vec{x}_i) = y_i],$$

где $[\cdot]$ – нотация (скобка) Айверсона:

$$[P] = \begin{cases} 1, \text{если } P - \text{истинно} \\ 0, \text{если } P - \text{ложно} \end{cases}$$

- Доля неправильных ответов:

$$Q(a, X) = \frac{1}{l}\sum_{i=1}^{l}[a(\vec{x}_i) \neq y_i] = \frac{1}{l}\sum_{i=1}^{l}[sign\langle\vec{w}, \vec{x}_i\rangle \neq y_i] \to \min_{\vec{w}}$$

– дискретная функция

Функционал ошибки

- Модифицированный вариант:

$$Q(a, X) = \frac{1}{l} \sum_{i=1}^{l} [y_i \langle \vec{w}, \vec{x}_i \rangle < 0] \to \min_{\vec{w}}$$

- Величина $M = y_i \langle \vec{w}, \vec{x}_i \rangle$ называется *отступом* (margin)
- Знак отступа говорит о корректности ответа классификатора:
  - положительный отступ $\to$ правильный ответ
  - отрицательный отступ $\to$ неправильный ответ
- Абсолютная величина отступа характеризует степень уверенности классификатора в своём ответе

Функция потерь

- Функция потерь (пороговая):

$$L(M) = [M < 0]$$

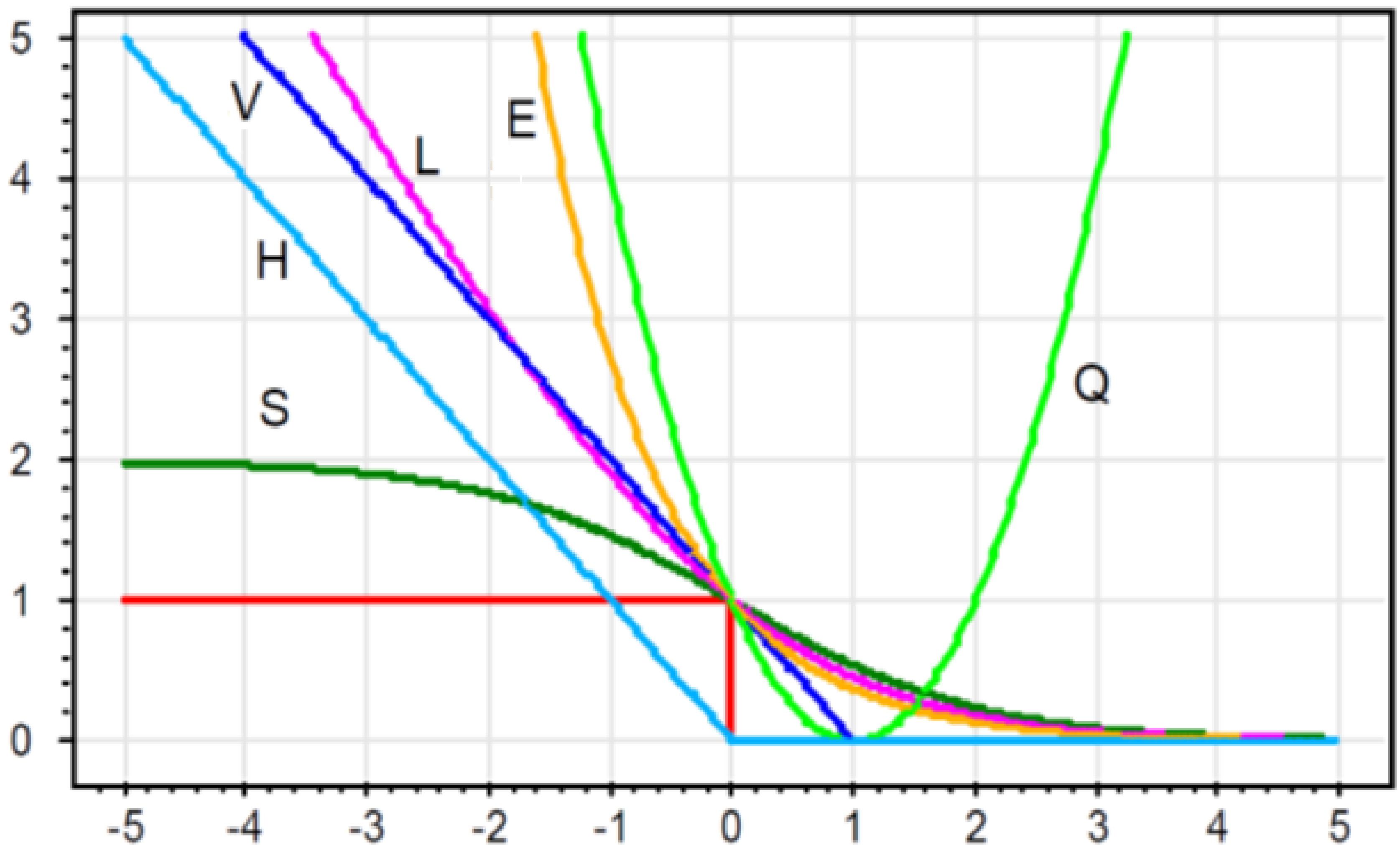
- Если оценить эту функцию сверху:

$$L(M) \leq \tilde{L}(M),$$

то можно получить верхнюю оценку для функционала ошибки:

$$Q(a, X) \leq \frac{1}{l} \sum_{i=1}^{l} \tilde{L}(y_i \langle \vec{w}, \vec{x}_i \rangle) \to \min_{\vec{w}}$$

Функция потерь



Функция потерь

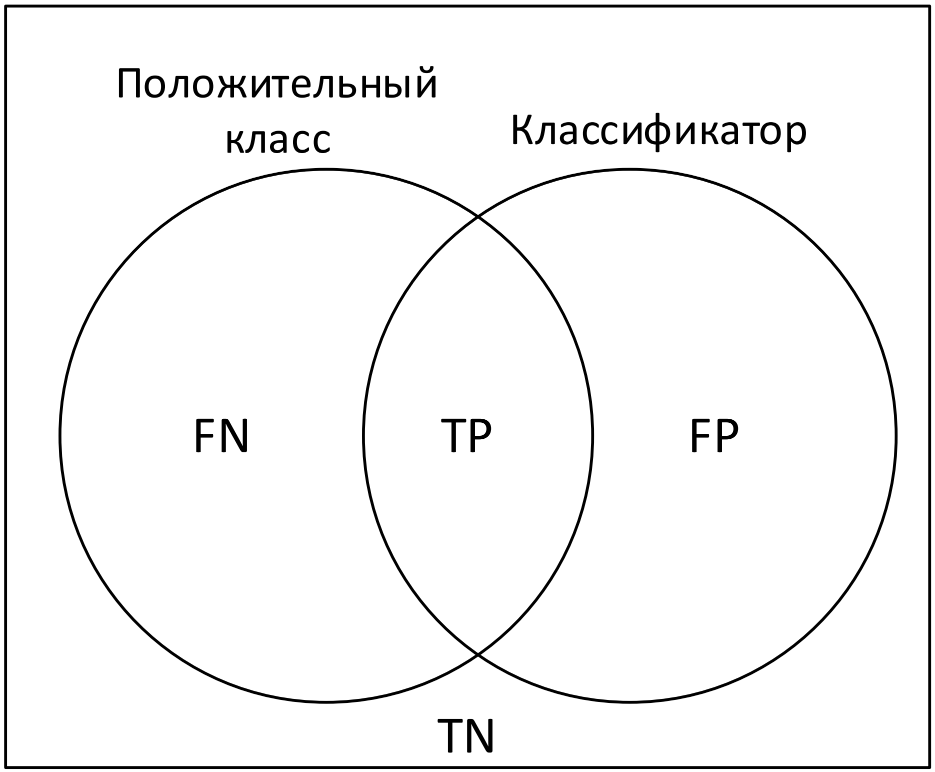
1. $\tilde{L}(M) = log(1 + e^{-M})$ – логистическая функция потерь (L)
2. $\tilde{L}(M) = (1 - M)_+ = max(0{,}1 - M)$ – кусочно-линейная функция потерь (hinge loss) (метод опорных векторов) (V)
3. $\tilde{L}(M) = (-M)_+ = max(0, -M)$ – кусочно-линейная функция потерь (персептрон Розенблатта) (H)
4. $\tilde{L}(M) = e^{-M}$ – экспоненциальная функция потерь (AdaBoost) (E)
5. $\tilde{L}(M) = \frac{2}{(1+e^{M})}$ – сигмоидная функция потерь (S)
6. $\tilde{L}(M) = (1 - M)^2$ – квадратичная функция потерь (Q)

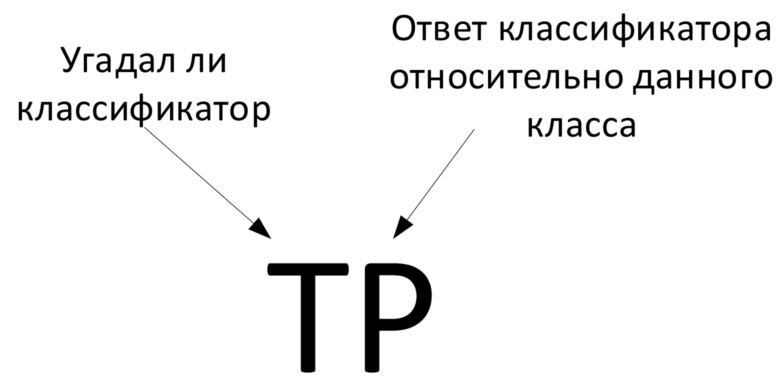
Метрики качества классификации

Матрица ошибок (confusion matrix):

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Оценка классификатора</td></tr>
<tr><td>$a(x) = +1$
(Positive)</td><td>$a(x) = -1$
(Negative)</td></tr>
<tr><td rowspan="2">Истинные ответы</td><td>$y = +1$</td><td>TP
(True Positive)</td><td>FN
(False Negative)</td></tr>
<tr><td>$y = -1$</td><td>FP
(False Positive)</td><td>TN
(True Negative)</td></tr>
</table>

Метрики качества классификации





Метрики качества классификации

- Accuracy (Правильность):

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

- Precision (Точность):

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

- Recall (Полнота):

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

- F1-measure (F1-score, F1-мера):

$$F1 = \frac{2PR}{P + R}$$

Метрики качества классификации

- Микроусреднение (micro-averaging):

$$P^{micro} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n}(TP_i + FP_i)}, \qquad R^{micro} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n}(TP_i + FN_i)},$$

$$F_1^{micro} = \frac{2P^{micro}R^{micro}}{P^{micro} + R^{micro}}$$

- Макроусреднение (macro-averaging):

$$P^{macro} = \frac{\sum_{i=1}^{n} P_i}{n}, \qquad R^{macro} = \frac{\sum_{i=1}^{n} R_i}{n},$$

$$F_1^{macro} = \frac{\sum_{i=1}^{n} F_{1i}}{n}$$