SCHOOL
OF ENERGY
& POWER ENGINEERING

Обучение классификаторов

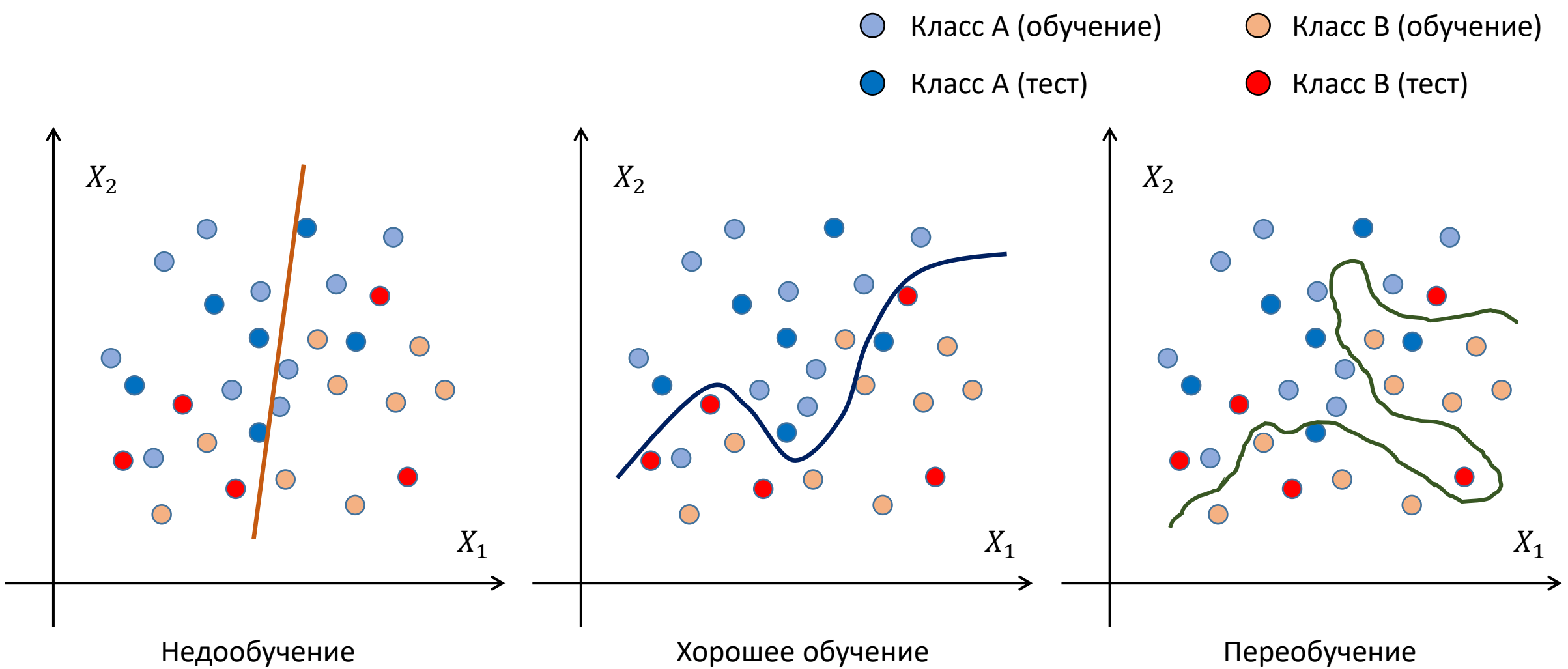
Сергей Владимирович Аксёнов,

Доцент отделения информационных технологий ИШИТР,

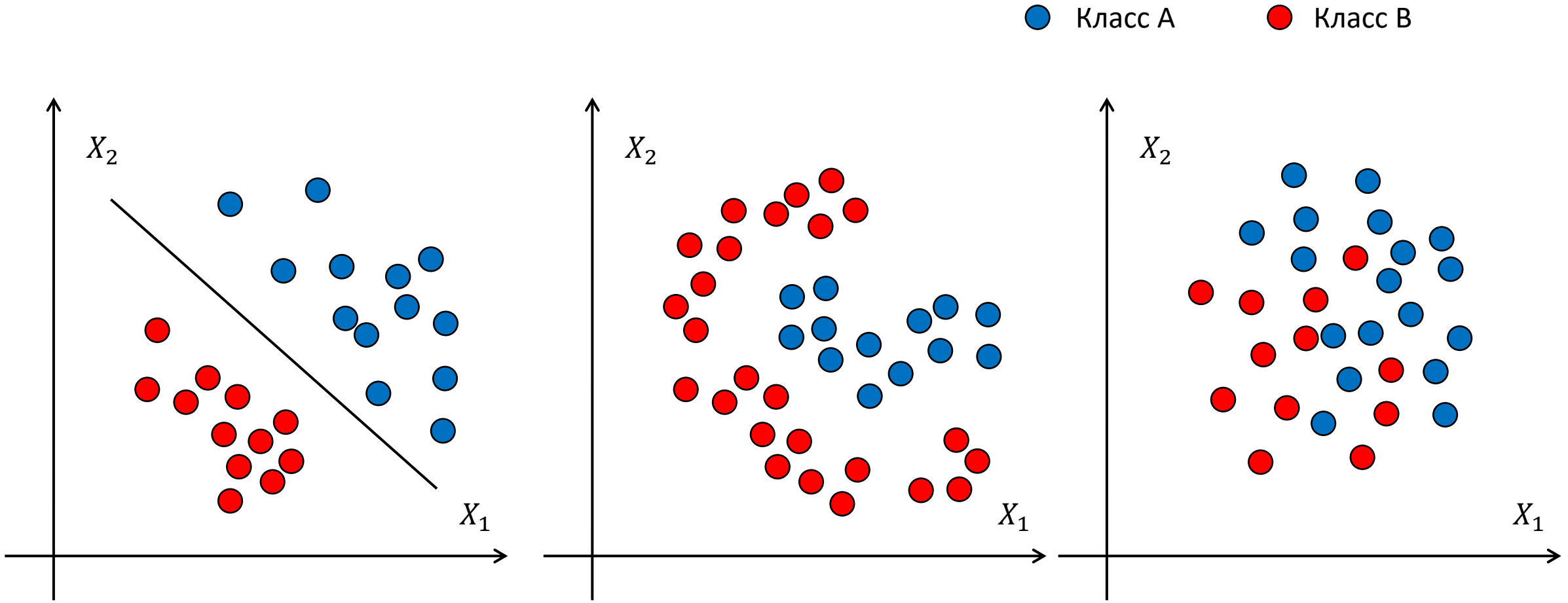
Томский политехнический университет

Томск-2023

Плохое и хорошее обучение



Линейно разделимые и линейно неразделимые классы



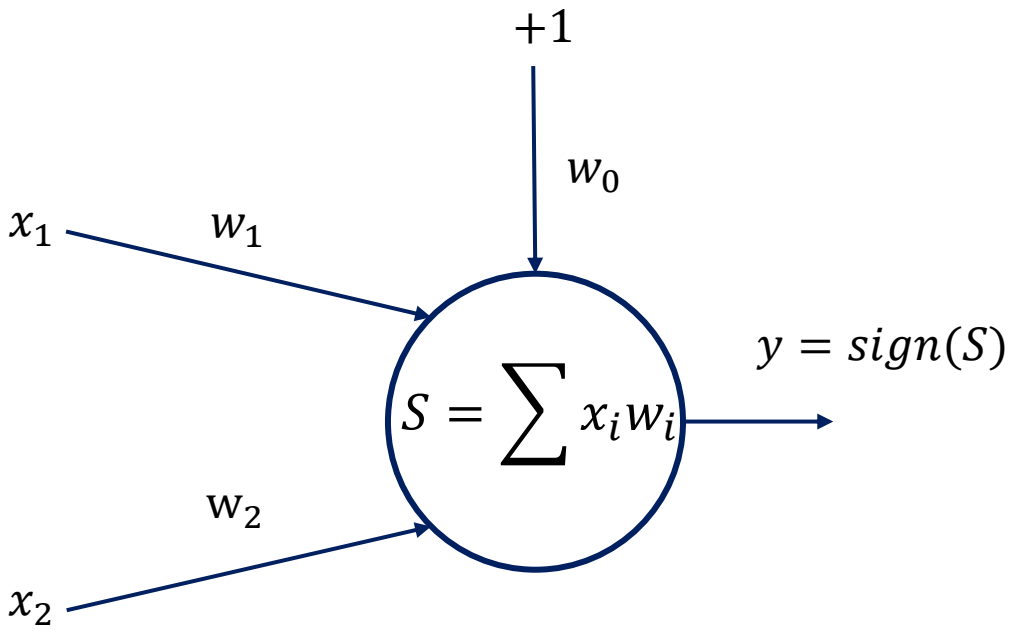
Линейно разделимые классы

Линейно неразделимые классы

Линейно неразделимые классы

Бинарный линейный классификатор-1

Вектор признаков: $x = (x_1, x_2, x_3, \dots, x_N)$



Положительный класс ($\hat{y}$=+1)

Отрицательный класс ($\hat{y}$=-1)

Выход модели: $\hat{y} = \hat{y}(x, w) = sign\left(w_0 + \sum_i^N w_i x_i\right) = sign(w^T x)$

Бинарный линейный классификатор-2

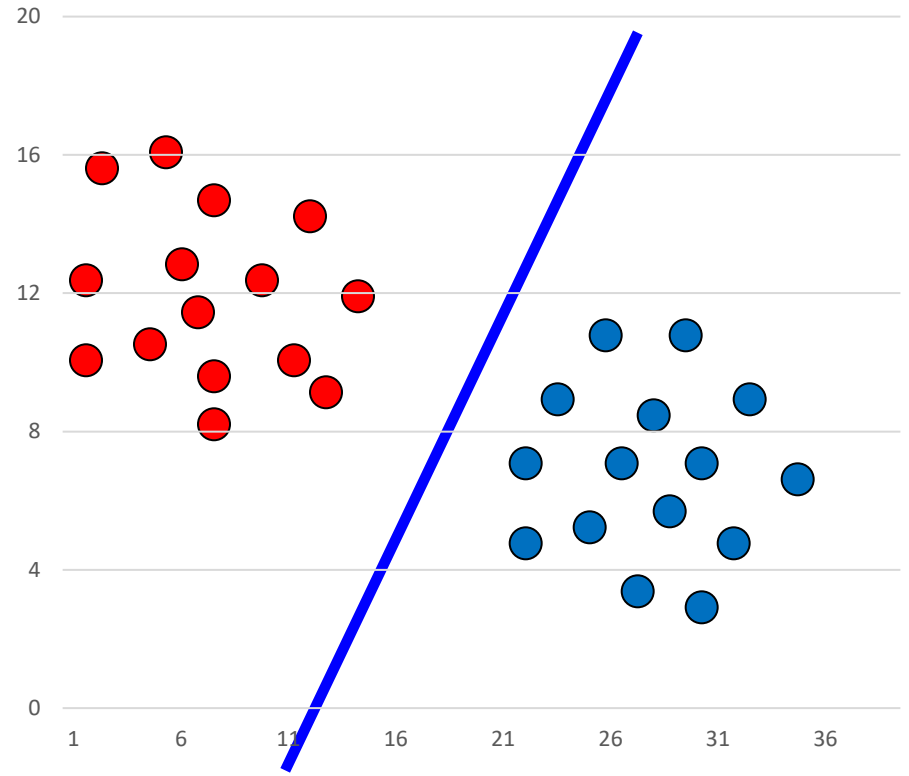
Результат обучения: входной вектор относится либо к положительному ($\hat{y}$=+1), либо отрицательному ($\hat{y}$=-1) классу

Вектор признаков:

$$x = (x_1, x_2, x_3, \dots, x_N)$$

Выход модели:

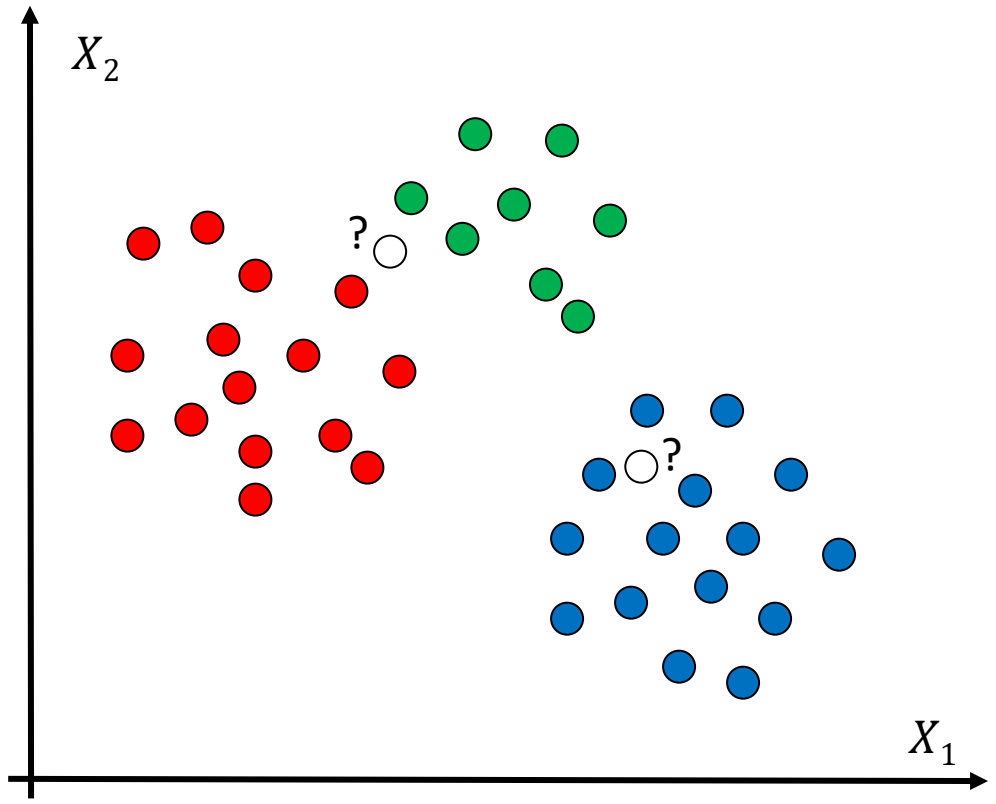
$$\hat{y} = \hat{y}(x, w) = sign\left(w_0 + \sum_i^N w_i x_i\right) = sign(w^T x)$$



К-ближайших соседей

Нахождение набора объектов, чьи признаки близки к тестирующему примеру.
К – число соседей = 1, 3, 5

Приведение признаков к одинаковой шкале.
Требование по хранению всей выборки.



Расстояния

$$D = \sqrt{\sum_{i=1}^{n} (Q_i - P_i)^2}$$ Эвклидово расстояние

$$D = \sum_{i=1}^{n} \frac{|Q_i - P_i|}{|Q_i| + |P_i|}$$ Расстояние Канберры

$$D = \sum_{i=1}^{n} |Q_i - P_i|$$ Расстояние Манхэттена

$$D = \frac{1}{n} \sum_{i=1}^{n} (Q_i - P_i)^2$$ MSE Расстояние

Метод опорных векторов (Support Vector Machine)

Результат обучения: максимизация зазора (расстояния между разделяющей гиперплоскостью и самыми близкими к этой плоскости тренировочными образцами

«Положительная» гиперплоскость: $w^T x = +1$

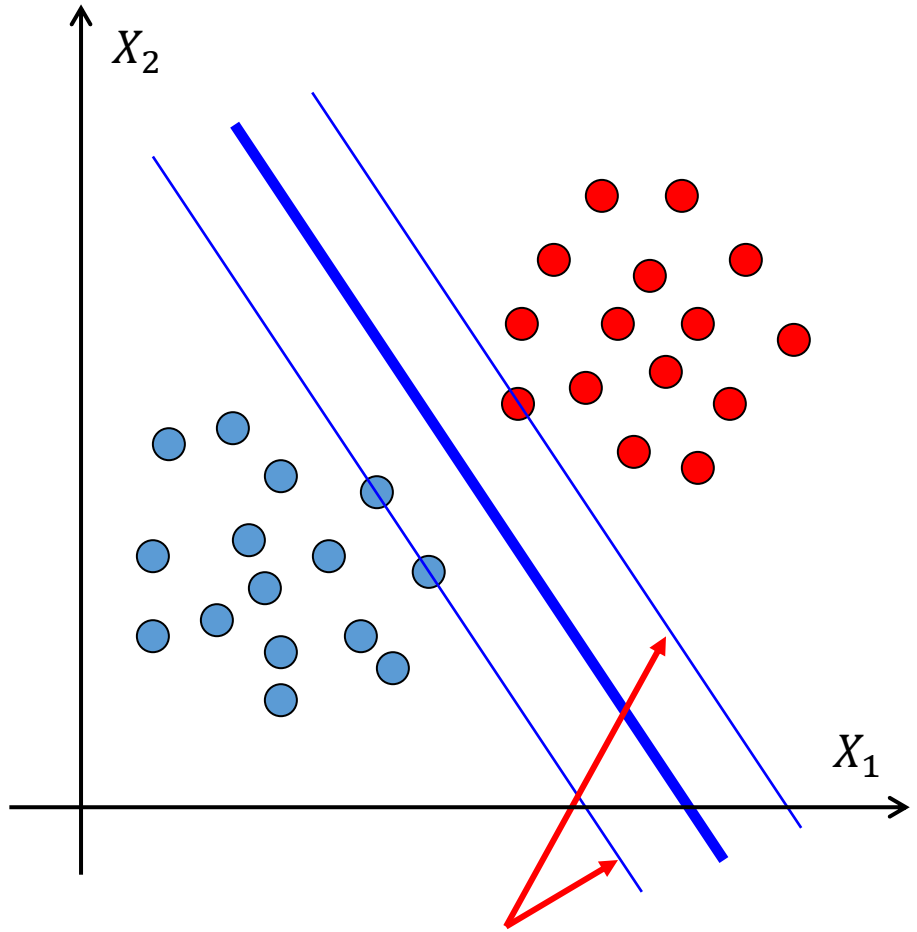
«Отрицательная» гиперплоскость: $w^T x = -1$

Граница решения: $w^T x = 0$

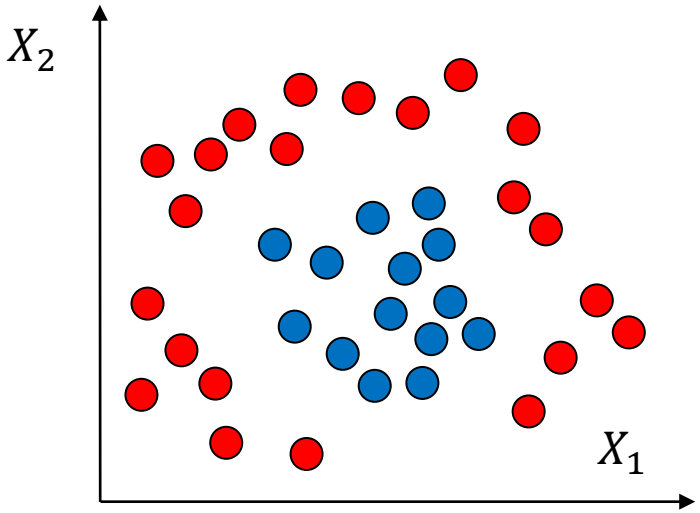
Целевая функция SVM: $\frac{2}{\|w\|} \rightarrow max$

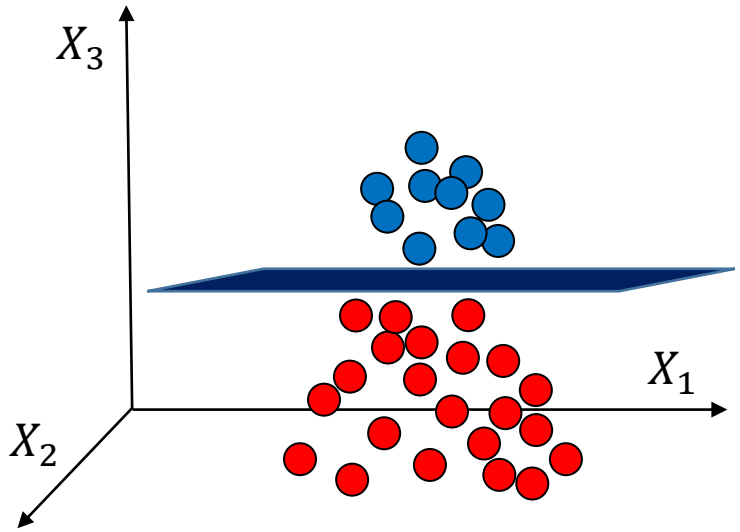
При ограничениях: $w_0 + w^T x^i \geq +1, \text{если } y^i = 1$

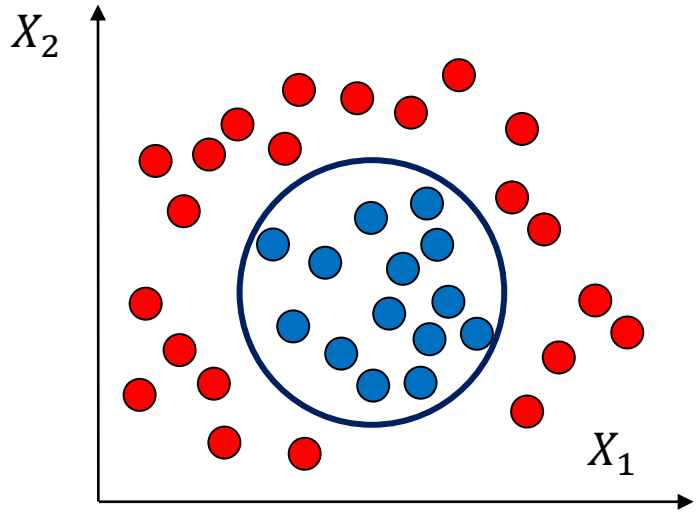
$w_0 + w^T x^i < -1, \text{если } y^i = -1$



Ядерный трюк SVM







$$X_3 = \phi(X_1, X_2)$$

Наиболее популярные ядра

Radial-basis function (RBF) kernel: $k(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2\sigma^2}\right)$ $d(x_i, x_j) = \|x_i - x_j\|$

Matérn kernel: $k(x_i, x_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{\sigma}d(x_i, x_j)\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}}{\sigma}d(x_i, x_j)\right)$

Rational quadratic kernel: $k(x_i, x_j) = \left(1 + \frac{d(x_i, x_j)^2}{2\alpha\sigma^2}\right)^{-\alpha}$

Exp-Sine-Squared kernel: $k(x_i, x_j) = \exp\left(-\frac{2sin^2\left(\pi d(x_i, x_j)/p\right)}{\sigma^2}\right)$

Dot-Product kernel: $k(x_i, x_j) = \sigma_0^2 + x_i \cdot x_j$

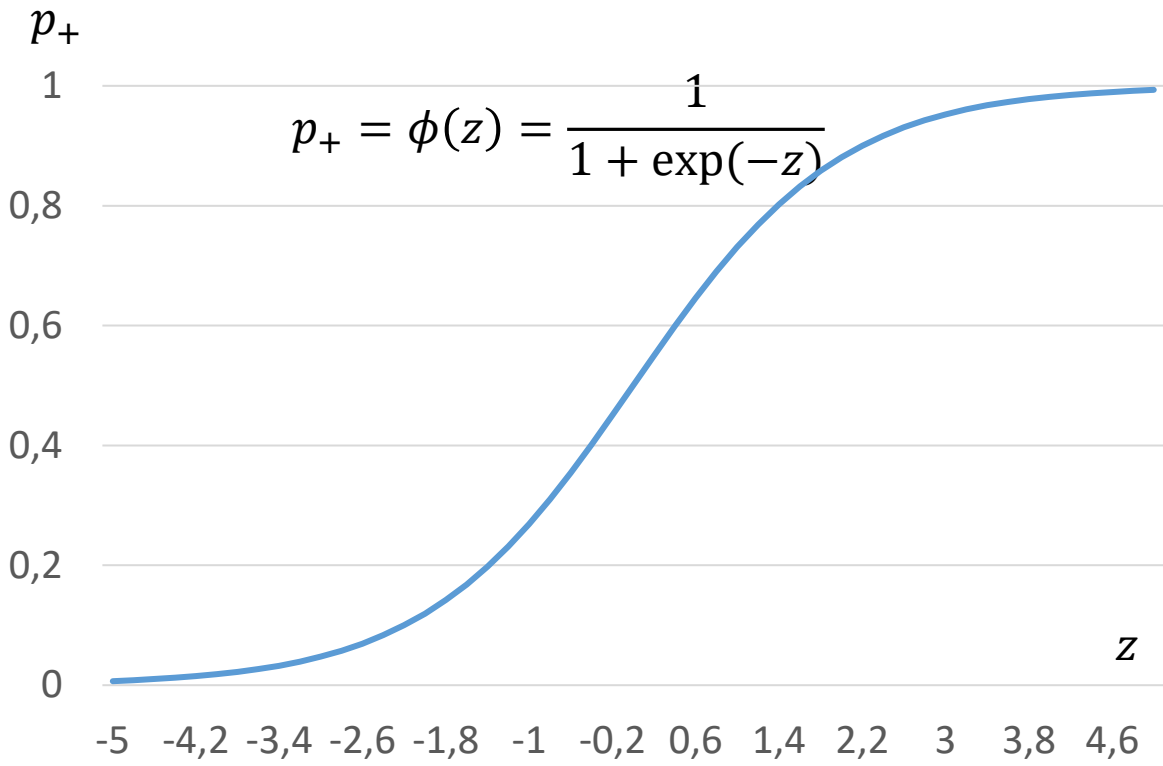
Логистическая регрессия

Прогнозируют вероятность p_+ отнесения примера x r к классу +1

$$z = \sum_i^N w_i x_i$$

Функция стоимости

$$J(w) = \frac{1}{2}\sum_i \left(\phi\left(z^{(i)}\right) - y^{(i)}\right)^2$$



Функция правдоподобия

$$L(w) = P(y|x;w) = \prod_{i=1}^{n} P\left(y^{(i)}|x^{(i)};w\right) = \prod_{i=1}^{n} \left(\phi\left(z^{(i)}\right)\right)^{y^{(i)}} \cdot \left(1-\phi\left(z^{(i)}\right)\right)^{1-y^{(i)}}$$

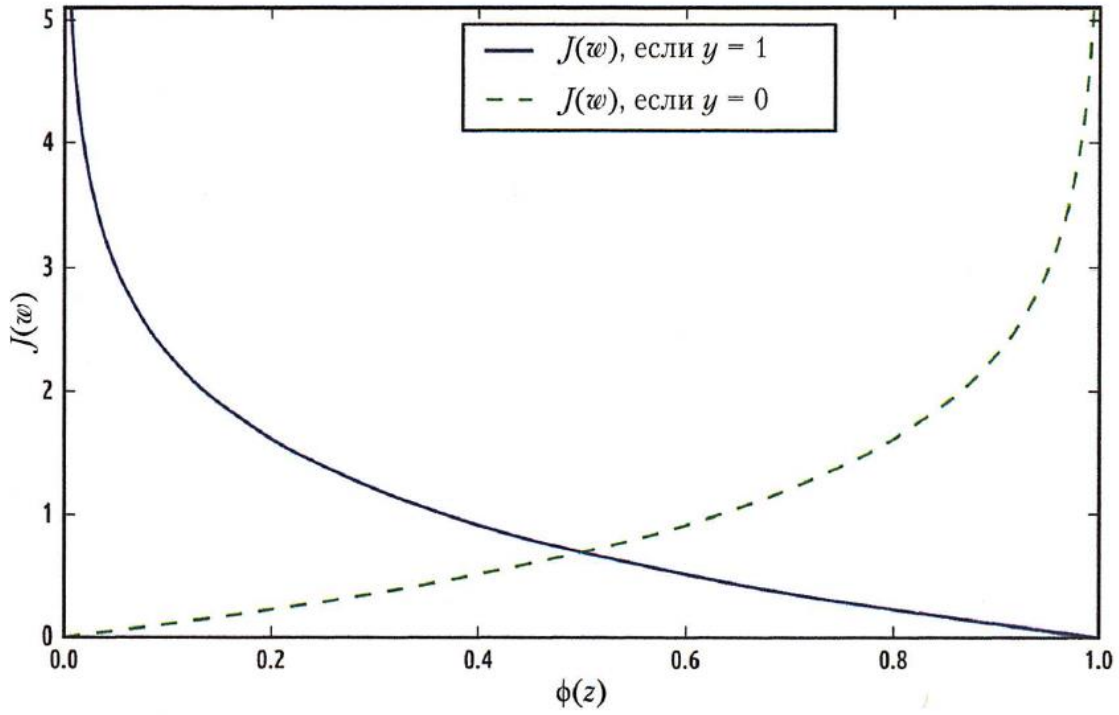
Логарифмическая функция правдоподобия

$$l(w) = logL(w) = \sum_{i=1}^{n} \left[y^{(i)} log\left(\phi\left(z^{(i)}\right)\right) + \left(1-y^{(i)}\right) log\left(1-\phi\left(z^{(i)}\right)\right)\right]$$

Логистическая регрессия: обучение

Функция, использующаяся для поиска параметров модели

$$J(w) = \sum_{i=1}^{n} \left[-y^{(i)} log\left(\phi\left(z^{(i)}\right)\right) - \left(1 - y^{(i)}\right) log\left(1 - \phi\left(z^{(i)}\right)\right) \right]$$



Регуляризация

L2 - регуляризация $\frac{\lambda}{2}\|w\|_2^2 = \frac{\lambda}{2}\sum_{j=1}^{m} w_j^2$

Новая функция, учитывающая штрафы $J(w) = \sum_{i=1}^{n}\left[-y^{(i)} log\left(\phi\left(z^{(i)}\right)\right) - \left(1 - y^{(i)}\right) log\left(1 - \phi\left(z^{(i)}\right)\right)\right] + \frac{\lambda}{2}\|w\|_2^2$

Дерево решений

Разбиение данных на подмножества, приводящему к самому большому приросту информации (получению однородных регионов решения)

Функция прироста информации:

$$IG\left(D_p, f\right) = I\left(D_p\right) - \frac{N_{left}}{N_p} I\left(D_{left}\right) - \frac{N_{right}}{N_p} I\left(D_{right}\right)$$

Меры неоднородности

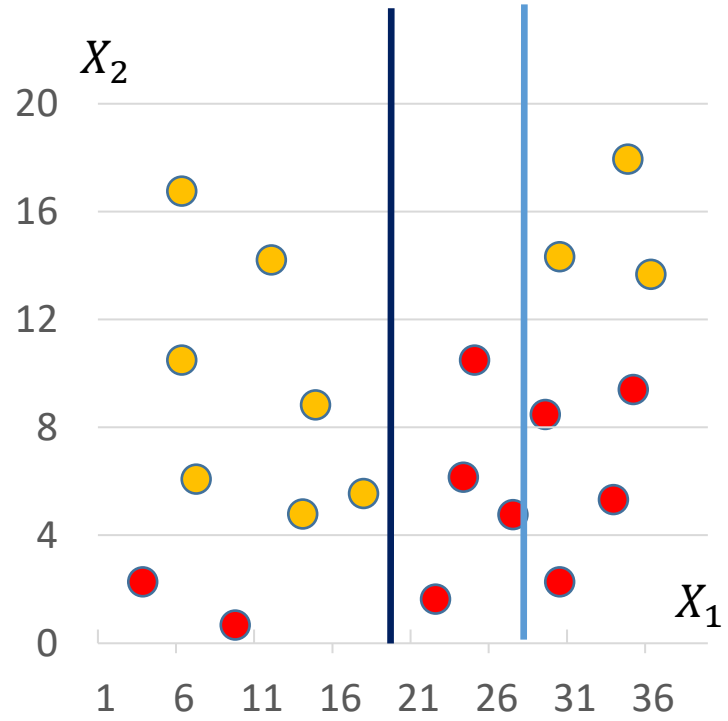
Энтропия: $I_G(t) = 1 - \sum_{i=1}^{c} p(i|t)^2$

Мера неопределенности Джини: $I_H(t) = -\sum_{i=1}^{c} p(i|t) log_2 p(i|t)$

Ошибка классификации: $I_E(t) = 1 - \max(p(i|t))$

$p(i|t)$ - доля образцов, принадлежащая классу i для узла t

Дерево решений: пример-1



$$IG(D_p, f) = I(D_p) - \frac{N_{left}}{N_p} I(D_{left}) - \frac{N_{right}}{N_p} I(D_{right})$$

В качестве критерия взята ошибка классификации:

$$I_E(t) = 1 - \max(p(i|t))$$

Неоднородность корневого узла:

$$I(D_0) = 1 - \max\left(\frac{10}{20}, \frac{10}{20}\right) = 1 - 0.5 = 0.5$$

Для расщепления $x_1 = 20$: $IG(D_0, x_1 = 20) = 0.5 - \frac{9}{20}\left(1 - \frac{7}{9}\right) - \frac{11}{20}\left(1 - \frac{8}{11}\right) = 0.25$

Для расщепления $x_1 = 28$: $IG(D_0, x_1 = 28) = 0.5 - \frac{13}{20}\left(1 - \frac{7}{13}\right) - \frac{7}{20}\left(1 - \frac{4}{7}\right) = 0.05$

Дерево решений: пример-2