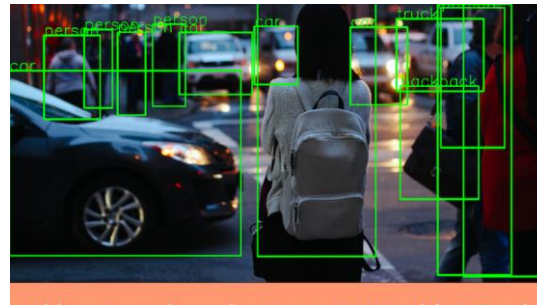
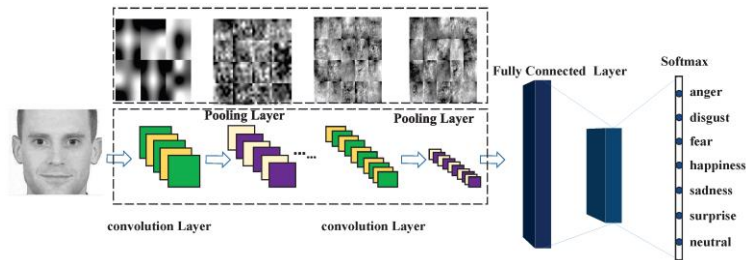


Основы искусственного интеллекта



Лекция

Обучение с учителем: Логистическая регрессия

к.ф.-м.н., доцент кафедры ИСиЦТ
Корнаева Е.П.

Обучение с учителем: Логистическая регрессия

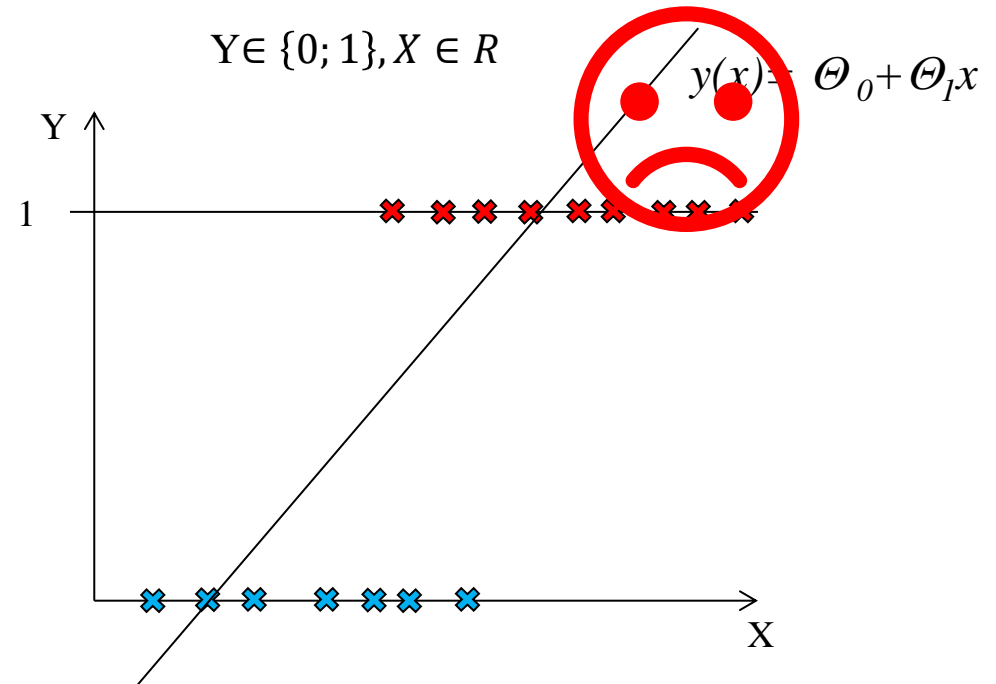
Постановка задачи классификации

Имеются k классов, условно обозначенных $y=0, y=1, \dots, y=k-1$ и объекты, обладающие m признаками X_i . Для n объектов известно к каким классам они относятся (размеченная выборка). Требуется на основании обучающей выборки отнести новые объекты с X_i признаками к конкретному классу.

Пример двухклассовой бинарной классификации

X	Y
x_1	1
x_2	0
	...
x_{n-1}	1
x_n	1

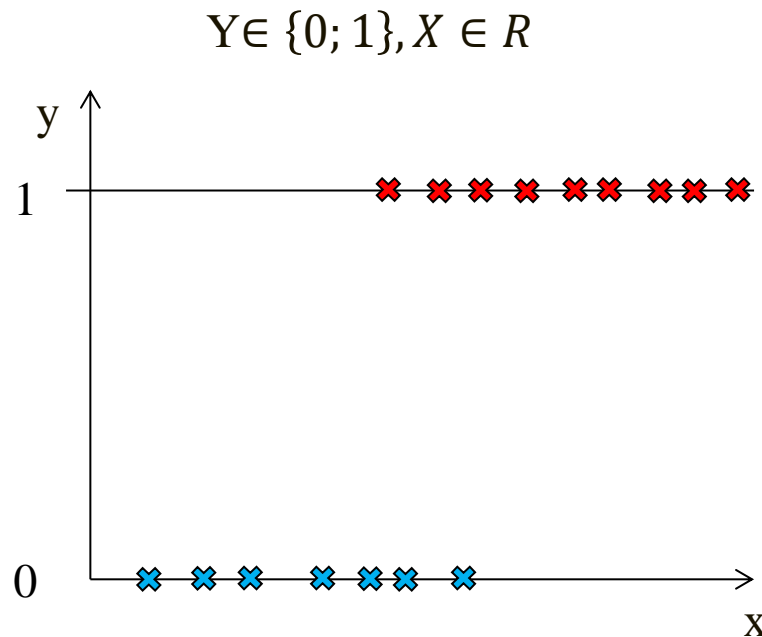
$X \in R, Y \in \{0; 1\}$



Обучение с учителем: Логистическая регрессия

Постановка задачи классической регрессии

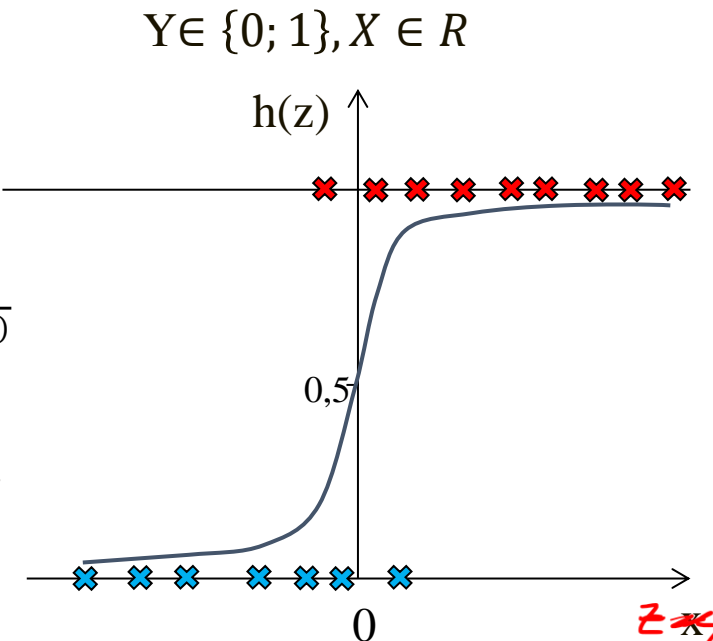
Логистическая регрессия - [статистическая](#) модель, используемая для прогнозирования вероятности возникновения некоторого события путём его сравнения с [логистической кривой](#) [wikipedia.org]



$$h_{\theta}(z) = \frac{1}{1 + e^{-z(x)}}$$

$$z(x) = \theta_0 + \theta_1 x$$

→



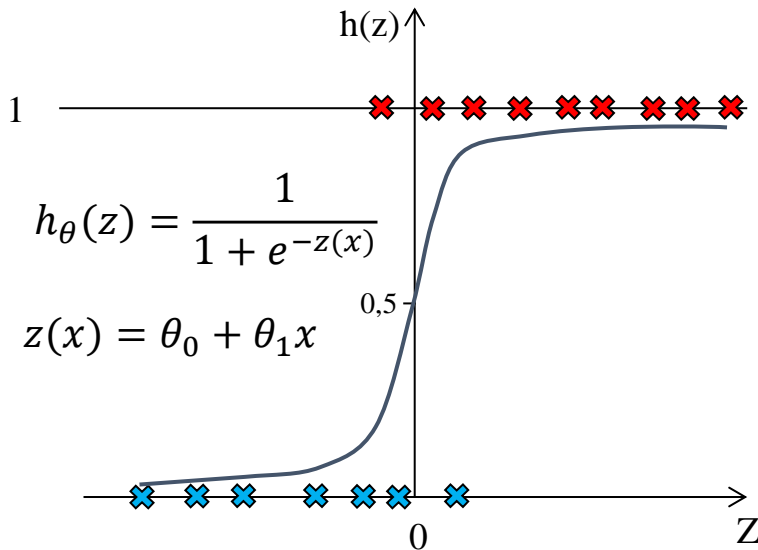
Обучение с учителем: Логистическая регрессия

Постановка задачи классической регрессии

Данные выборки представлены множеством пар: $\{(X_i, y_i)\}, i = \overline{1, n}, Y = \{0, 1\}$

Аппроксимация вероятности P принадлежности к классу "1" как функцию признаков X

$Y \in \{0; 1\}, X \in R$



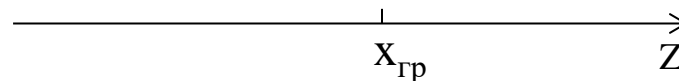
$h(z) = P(y = 1) \in [0; 1]$ – вероятность принадлежности к классу «1»:

- если $h_\theta(z) \geq 0.5$, то объект принадлежит классу $y=1$;
- если $h_\theta(z) < 0.5$, то объект принадлежит классу $y=0$.

$$h_\theta(z) = 0.5: z(x) = 0$$

или

- если $z(x) \geq 0$, то объект принадлежит классу $y=1$;
- если $z(x) < 0$, то объект принадлежит классу $y=0$.



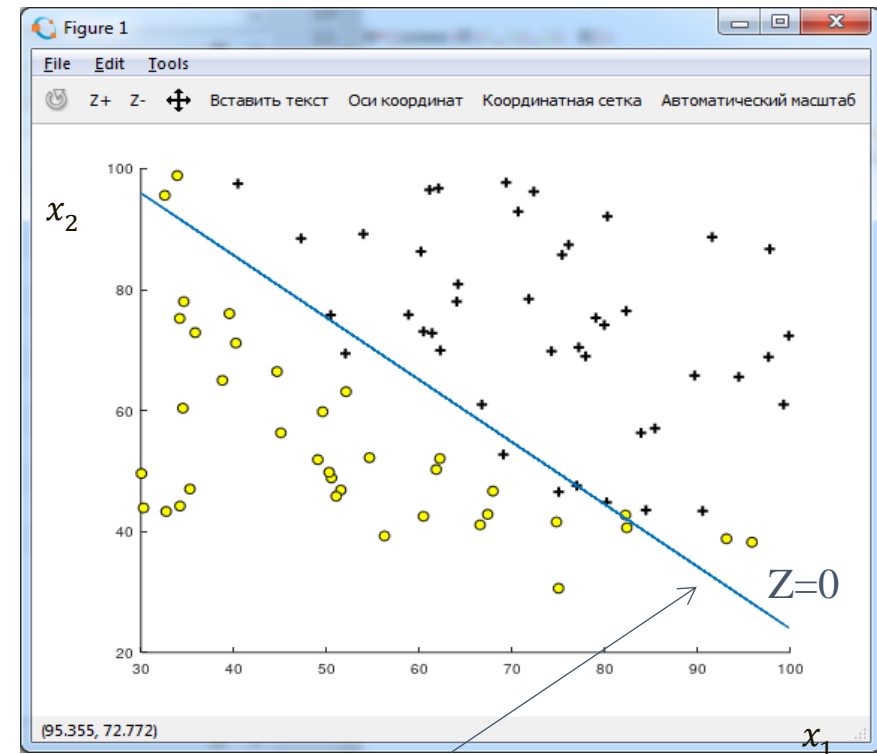
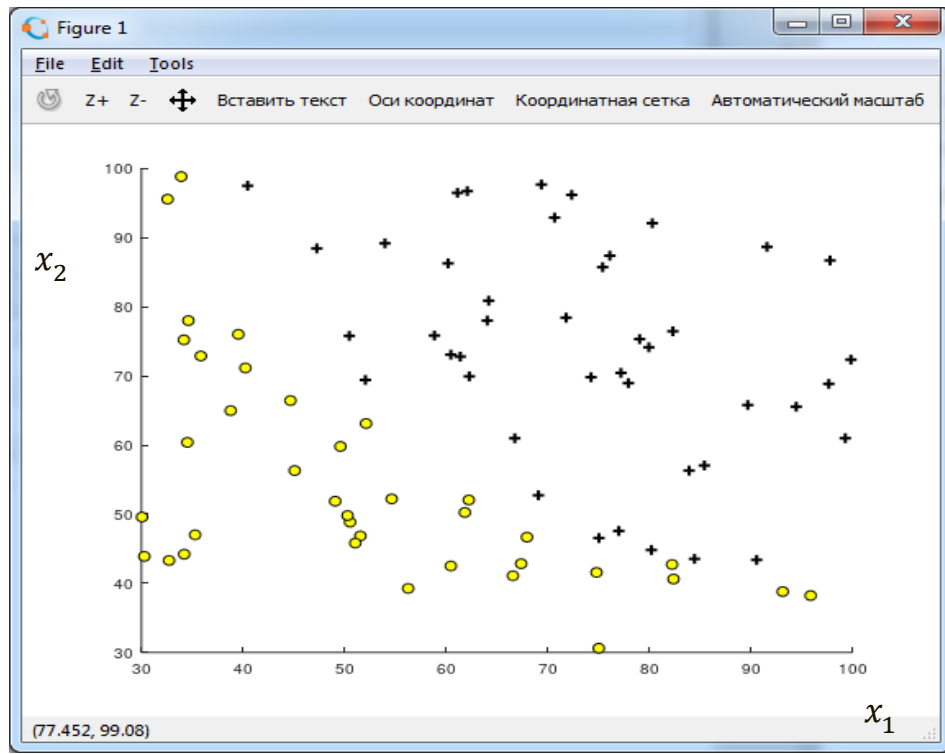
$$z(x) = 0: \theta_0 + \theta_1 x_{\text{gp}} = 0 \Rightarrow x_{\text{gp}} = -\frac{\theta_0}{\theta_1}$$

Обучение с учителем: Логистическая регрессия

Постановка задачи классической регрессии

Данные выборки представлены множеством пар: $\{(X_i, y_i)\}, i = \overline{1, n}, Y = \{0, 1\}$

Пример Двумерный случай: $X = [[X_1, X_2]]$



$$z(X) = 0,$$

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 = 0 \Rightarrow x_2(x_1) = -\frac{\theta_0}{\theta_2} - \frac{\theta_1}{\theta_2} x_1$$

Обучение с учителем: Логистическая регрессия

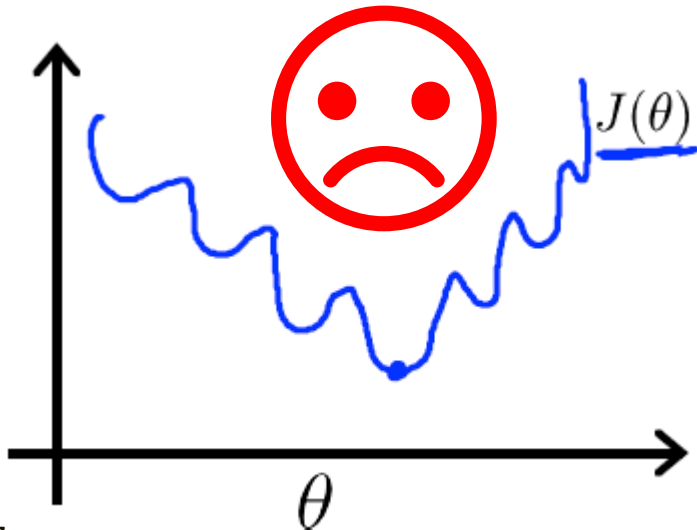
Логистическая регрессия. Определение параметров модели

$$h_{\theta}(x) = \frac{1}{1 + \exp(-z(x))}, \quad \theta = [\theta_i] - \text{параметры модели}$$

$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

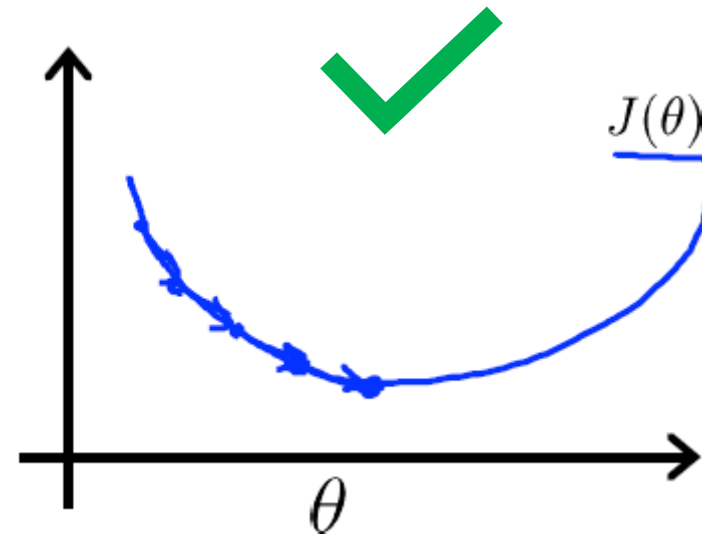
Задача безусловной оптимизации функции ошибки:

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n (h_{\theta}(x_i) - y_i)^2 \rightarrow \min$$



Cross Entropy loss

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_{\theta}(x_{ij})) + (1 - y_i) \ln(1 - h_{\theta}(x_{ij}))) \rightarrow \min$$



Логистическая регрессия. Определение параметров модели

Y — случ. величина: $y_i \in \{0; 1\}$

В n независимых испытаниях, в которых событие $y_i=1$ может появиться либо нет.

Вероятность p наступления события постоянна.

$$\text{sum}(y_i==1) = n_1$$

p_i - вероятность события $y_i=1$

$$\text{sum}(y_i==0) = n_2$$

$q_i = 1 - p_i$ - вероятность события $y_i=0$

$$P(y_i) = p_i^{y_i} (1 - p_i)^{1-y_i} = \begin{cases} p_i, & \text{если } y_i = 1 \\ 1 - p_i, & \text{если } y_i = 0 \end{cases}$$

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели

Y — случ. величина: $y_i \in \{0; 1\}$

В n независимых испытаниях, в которых событие $y_i=1$ может появиться либо нет.

Вероятность p наступления события постоянна.

$$\begin{array}{l|l} \text{sum}(y_i==1) = n_1 & \text{sum}(y_i==0) = n_2 \\ p_i = h(\theta, y_i) - \text{вероятность события } y_i=1 & q_i = 1 - p_i - \text{вероятность события } y_i=0 \end{array}$$

Функция максимального правдоподобия:

$$L(\theta, Y) = \prod_{i=1}^n P_i = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} = \prod_{i=1}^n h_i^{y_i} (1 - h_i)^{1-y_i} \xrightarrow{\theta} \max$$

Функция ошибки:

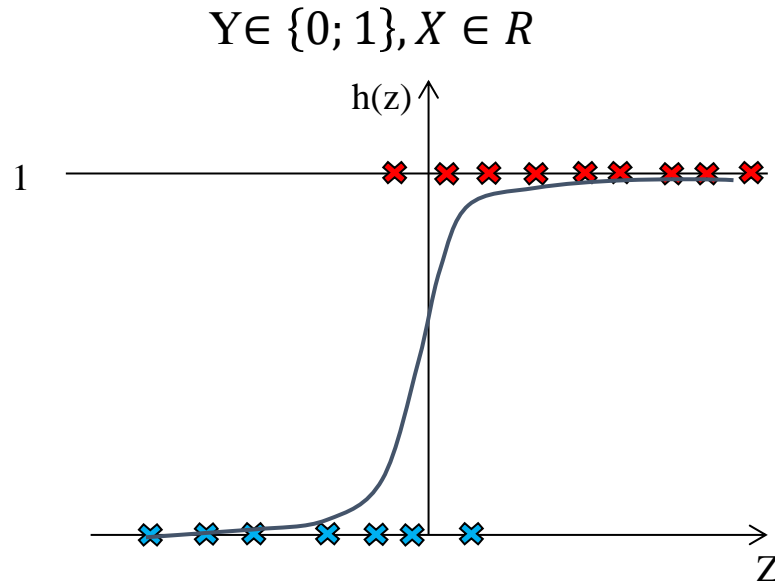
$$J(\theta) = -\ln(L(\theta, Y)) \xrightarrow{\theta} \min$$

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_i) + (1 - y_i) \ln(1 - h_i)) \rightarrow \min$$

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели

Функция ошибки:



$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_{\theta}(x)) + (1 - y_i) \ln(1 - h_{\theta}(x))) \rightarrow \min$$

$$h_{\theta}(x) = \frac{1}{1 + \exp(-z(x))} \quad \text{- логистическая функция (сигмоида)}$$

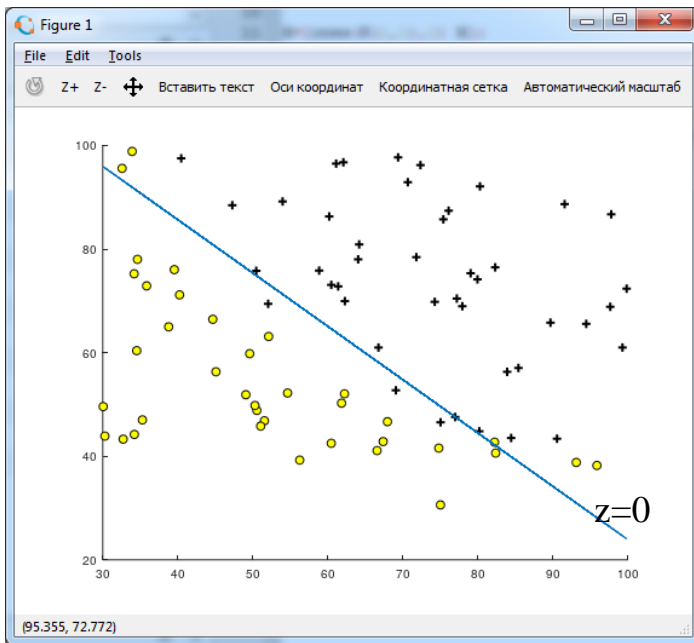
Примечание: в качестве $h_{\theta}(x)$ могут быть другие функции из семейства сигмоидных

Логистическая регрессия. Определение параметров модели на примере сигмоиды

I. Линейная граница между классами

Пример Линейная граница для 2-d случая

$$z(X) = \theta_0 + \theta_1 X_1 + \theta_2 X_2$$



Общий случай Линейная граница для многомерного случая:

$$z(X) = \theta_0 + \theta_1 X_1 + \dots + \theta_m X_m \stackrel{\text{red}}{=} \sum_{i=0}^m \theta_i X_i^*,$$

где $X = \llbracket X_i \rrbracket$ - факторное пространство;
 $X_0 = \llbracket [1 \dots 1] \rrbracket^T$.

Логистическая регрессия:

$$h_{\theta}(X) = \frac{1}{1 + \exp(-\theta_i X_i)}.$$

* Используется правило Эйнштейна

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

I. Линейная граница между классами

Решается задача минимизации функции ошибки:

$$J(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_{\boldsymbol{\theta}}(x_{ij})) + (1 - y_i) \ln(1 - h_{\boldsymbol{\theta}}(x_{ij}))) \rightarrow_{\boldsymbol{\theta}} \min$$

$$h_{\boldsymbol{\theta}}(\mathbf{X}) = \frac{1}{1 + \exp(-z(\mathbf{X}))}, \quad z(\mathbf{X}) = \sum_{i=0}^m \theta_i X_i$$

Необходимое условие экстремума:

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \theta_j} = \frac{1}{n} \sum_{i=1}^n (h_{\boldsymbol{\theta}}(x_{ij}) - y_i) x_{ij} = 0, \quad j = \overline{0, m}$$

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

I. Линейная граница между классами

Матричная форма записи:

	X_0	X_1	...	X_m	Y
1	1	x_{11}		x_{1m}	y_1
2	1	x_{21}		x_{2m}	y_2
...			...		
n	1	x_{n1}		x_{nm}	y_n

X

$$\underline{X}_{[n \times (m+1)]}, \quad Y_{[n \times 1]}, \quad \theta_{[m+1, 1]},$$

$$Z_{[n \times 1]} = Z(X), \quad H_{[n \times 1]} = \text{sigm}(Z)$$

В покомпонентном виде:

$$z_i = \sum_{j=0}^m \theta_j x_{ij}, i = 1..n$$

$$h_{\theta_i} = \frac{1}{1 + \exp(-z_i)},$$

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_{\theta}(x_{ij})) + (1 - y_i) \ln(1 - h_{\theta}(x_{ij})))$$

В матричном виде:

$$Z = X\theta,$$

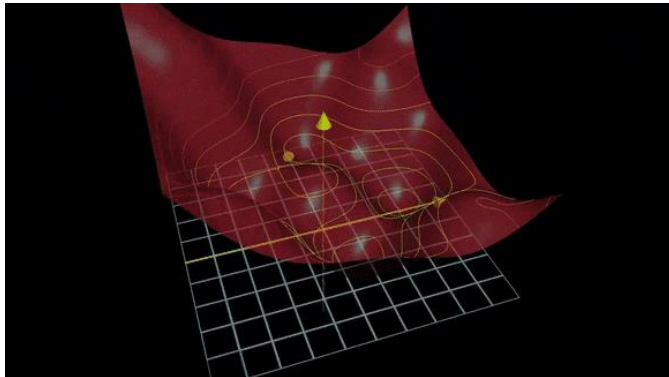
$$H = \text{sigm}(Z)$$

$$J(\theta) = -\frac{1}{n} (Y^T \ln H + (I - Y)^T \ln(I - H)) \xrightarrow{\theta} \min$$

где $I = (1)_{[n \times 1]}$

Логистическая регрессия. Определение параметров модели на примере сигмоиды

I. Линейная граница между классами. Численное решение методом градиентного спуска



Gradient descent intuition

Компоненты градиента:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{n} \sum_{i=1}^n (h_{\theta}(x_{ij}) - y_i) x_{ij} \quad \left| \quad \nabla J = \frac{1}{n} X^T (H - Y), \right.$$

Расчет коэффициентов на каждом шаге $l+1$:

$$\theta_j^{l+1} = \theta_j^l - \alpha \frac{\partial J}{\partial \theta_j} |_l \quad \left| \quad \boldsymbol{\theta}^{l+1} = \boldsymbol{\theta}^l - \alpha \nabla J |_l \right.$$

$\boldsymbol{\theta}^0 = (\theta_j^0)$ — начальное значение коэффициентов;

$\nabla J |_l = \left(\frac{\partial J}{\partial \theta_j} \right) |_l$ — компоненты градиента рассчитываются на каждом шаге $l+1$;

α — шаг градиентного спуска (скорость обучения).

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

I. Линейная граница между классами. Численное решение методом градиентного спуска

Компоненты градиента:

$$\nabla J = \frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{n} \sum_{i=1}^n (h_{\theta}(x_{ij}) - y_i) x_{ij} \quad \left| \quad \nabla J = \frac{1}{n} \mathbf{X}^T (\mathbf{H} - \mathbf{Y}), \right.$$

Расчет коэффициентов на каждом шаге l :

$$\theta_j^{l+1} = \theta_j^l - \lambda \frac{\partial J}{\partial \theta_j} |_l \quad \left| \quad \boldsymbol{\theta}^{l+1} = \boldsymbol{\theta}^l - \alpha \nabla J |_l \right.$$

Расчет функции ошибки на каждом шаге позволяет оценить процесс сходимости решения:

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_{\theta}(x_{ij})) + (1 - y_i) \ln(1 - h_{\theta}(x_{ij}))) \quad \left| \quad J = -\frac{1}{n} (\mathbf{Y}^T \ln \mathbf{H} + (\mathbf{I} - \mathbf{Y})^T \ln(\mathbf{I} - \mathbf{H})) \right.$$

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

II. Нелинейная граница между классами. Линеаризация

$z(\mathbf{X})$ – полином степени d .

Пример для случая $d=2$, $m=2$.

$z(\mathbf{X}) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_1 X_2 + \theta_4 X_1^2 + \theta_5 X_2^2$ пусть $X_3 = X_1 X_2$, $X_4 = (X_1)^2$, $X_5 = (X_2)^2$

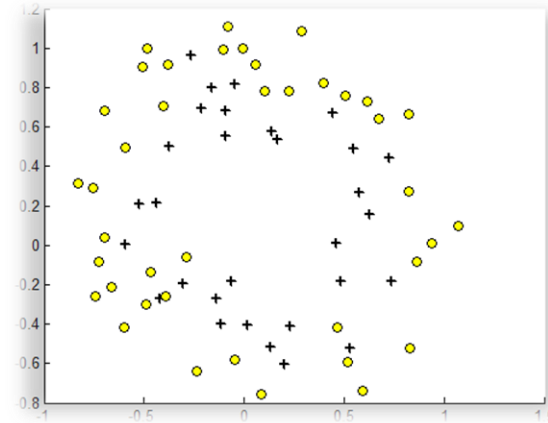
	X_0	X_1	X_2	X_3	X_4	X_5	\mathbf{Y}
1	1	x_{11}	x_{12}	$x_{11}x_{12}$	$(x_{11})^2$	$(x_{12})^2$	y_1
2	1	x_{21}	x_{22}	$x_{21}x_{22}$	$(x_{21})^2$	$(x_{22})^2$	y_2
...							
n	1	x_{n1}	x_{n2}	$x_{n1}x_{n2}$	$(x_{n1})^2$	$(x_{n2})^2$	y_n

$\mathbf{X} [n \times k]$

$$z(\mathbf{X}) = \sum_{i=0}^k \theta_i X_i = \mathbf{X}\boldsymbol{\theta},$$

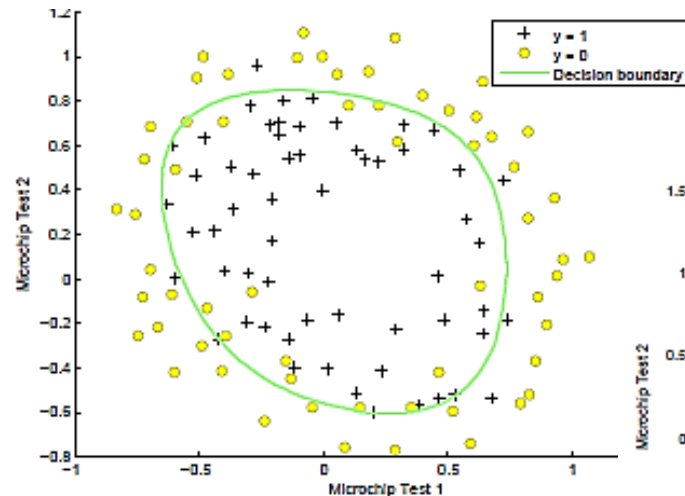
$$\boldsymbol{\theta} = (\theta_i)_{k \times 1}$$

$$J = -\frac{1}{n} (\mathbf{Y}^T \ln \mathbf{H} + (\mathbf{I} - \mathbf{Y})^T \ln(\mathbf{I} - \mathbf{H})) \rightarrow \min$$

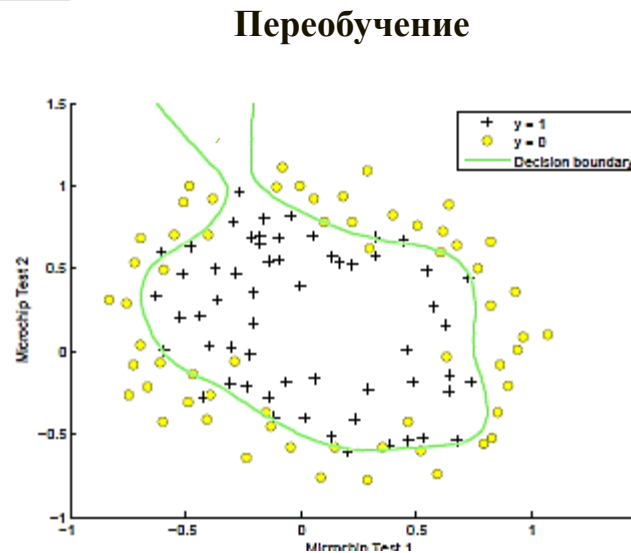


Логистическая регрессия. Определение параметров модели на примере сигмоиды

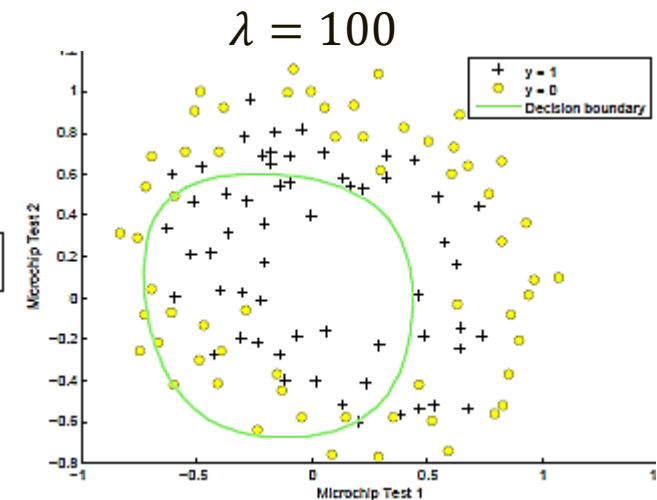
II. Нелинейная граница между классами. Регуляризация



$\lambda = 1$



$\lambda = 0$



Недообучение

L_2 – регуляризация:

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_{\theta}(x_{ij})) + (1 - y_i) \ln(1 - h_{\theta}(x_{ij}))) + \frac{\lambda}{2n} \sum_{j=1}^k \theta_j^2 \rightarrow \min$$

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

II. Нелинейная граница между классами. Регуляризация

Функция ошибки:

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_{\theta}(x_{i0})) + (1 - y_i) \ln(1 - h_{\theta}(x_{i0}))) + \frac{\lambda}{2n} \sum_{j=1}^k \theta_j^2 \rightarrow \min$$

Компоненты градиента функции ошибки:

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{n} \sum_{i=1}^n (h_{\theta}(x_{i0}) - y_i) x_{i0}$$

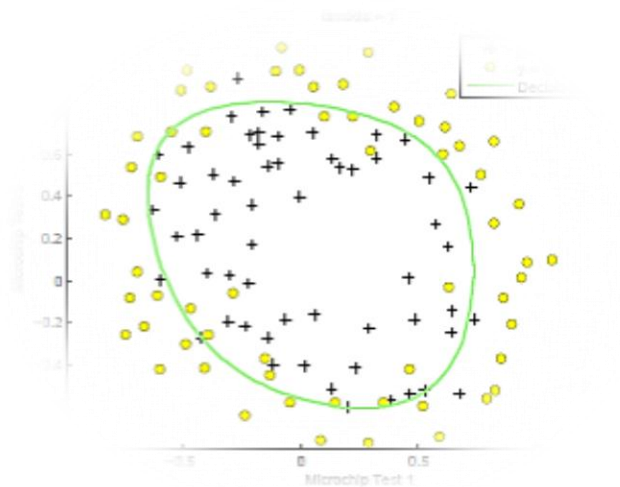
$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{n} \sum_{i=1}^n (h_{\theta}(x_{ij}) - y_i) x_{ij} + \frac{\lambda}{n} \theta_j, j = \overline{1, k}$$

$$\nabla J = \frac{1}{n} \mathbf{X}^T (\mathbf{h} - \mathbf{Y}) + \frac{\lambda}{n} \boldsymbol{\theta}$$

$$\nabla J_0 = \nabla J_0 - \frac{\lambda}{n} \theta_0$$

Логистическая регрессия. Определение параметров модели на примере сигмоиды

II. Регуляризация. Подбор гиперпараметров модели



- Модель: $h_{\theta}(x) = \frac{1}{1 + \exp(-z(x))}$, где $z(x)$ – полином степени d .

- Функция ошибки:

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_{\theta}(x_{ij})) + (1 - y_i) \ln(1 - h_{\theta}(x_{ij}))) + \frac{\lambda}{2n} \sum_{j=1}^k \theta_j^2 \rightarrow \min$$

На этапе *валидации* происходит выбор гиперпараметров, в данном случае (d, λ) :

- для каждого набора (d_i, λ_j) определяются параметры модели $\theta = \llbracket \theta_k \rrbracket$ (по обучающей выборке);
- для каждой модели рассчитывается ошибки J_{tr} и J_v (по обучающей и проверочной выборке):

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_{\theta}(x_{ij})) + (1 - y_i) \ln(1 - h_{\theta}(x_{ij})));$$

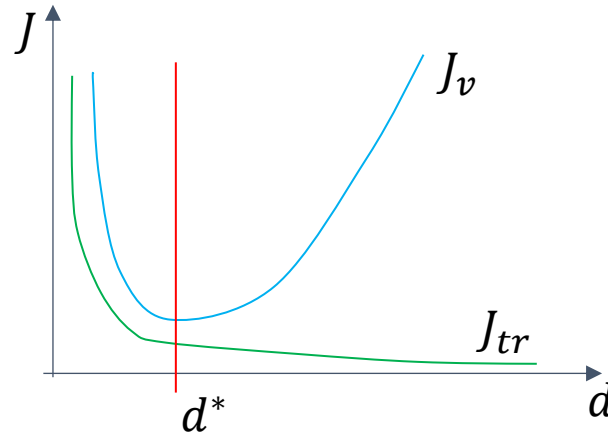
- по значениям J_{tr} и J_v выбирается (d_i, λ_j) .

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

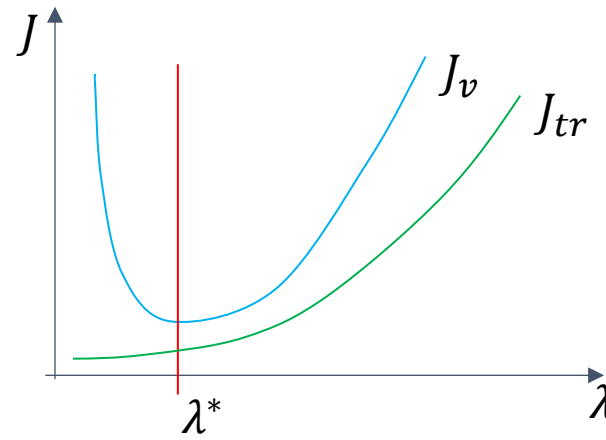
III. Регуляризация. Подбор гиперпараметров модели

$d \ll d^*$ - недообучение
(high bias)



$d \gg d^*$ - переобучение
(high variance)

$\lambda \ll \lambda^*$ - переобучение
(high variance)



$\lambda \gg \lambda^*$ - недообучение
(high bias)

Логистическая регрессия. Определение параметров модели на примере сигмоиды

III. Регуляризация. Подбор гиперпараметров модели

$$h_{\theta}(x) = \frac{1}{1 + \exp(-z(x))}, \text{ где } z(x) \text{ — полином степени } d.$$

- Выборка делится на три части: для обучения, валидации и тестирования;
- По обучающей выборке параметры модели $\theta = (\theta_i)_{k \times 1}$ определяются для каждой комбинации (d_i, λ_j) ;
- Выбираются оптимальные гиперпараметры (d^*, λ^*) , например сравниваются ошибки J_{tr}, J_v (рассчитанные по обучающей и проверочной выборке) для каждой комбинации (d_i, λ_j)
- Рассчитывается точность модели (см. далее) на тестовой выборке.

Обучение с учителем: Логистическая регрессия

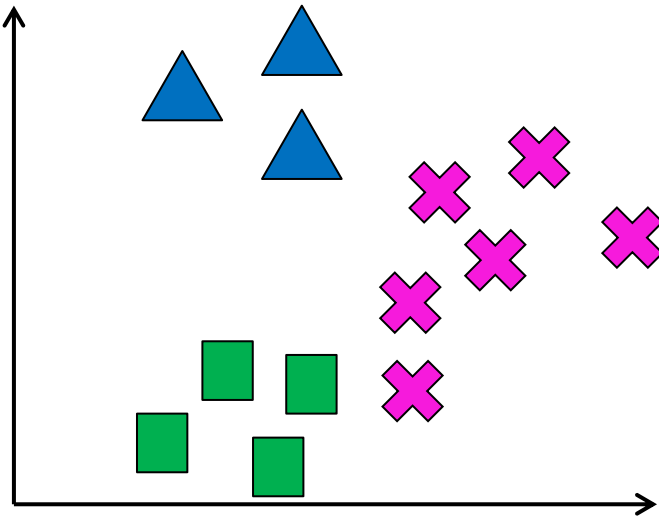
Многоклассовая классификация на примере метода «один против всех» (one-vs-all)

Метод «Один против всех»

Пример для трех классов.

$X \in R, Y \in \{1; 2; 3\}$

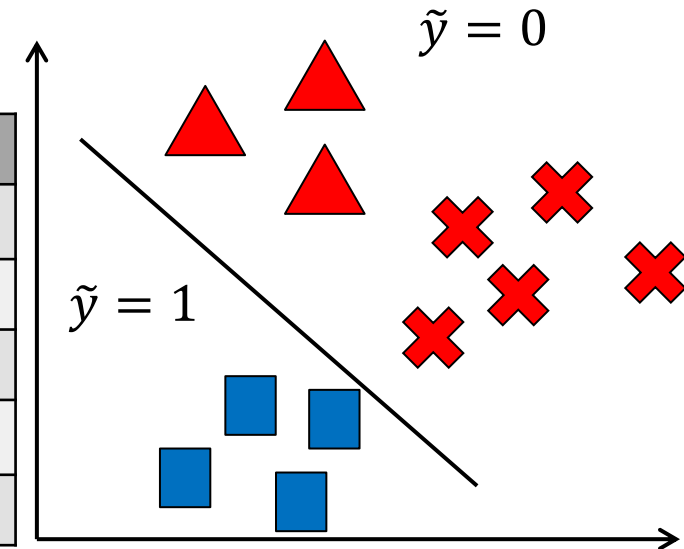
X_1	X_2	Y
x_{11}	x_{12}	1
x_{21}	x_{22}	3
		...
$x_{n-1,1}$	$x_{n-1,2}$	1
$x_{n,1}$	$x_{n,2}$	2



1 шаг «1 класс против всех»

X_1	X_2	Y
x_{11}	x_{12}	1
x_{21}	x_{22}	3
		...
$x_{n-1,1}$	$x_{n-1,2}$	1
$x_{n,1}$	$x_{n,2}$	2

\tilde{y}
1
0
...
1
0



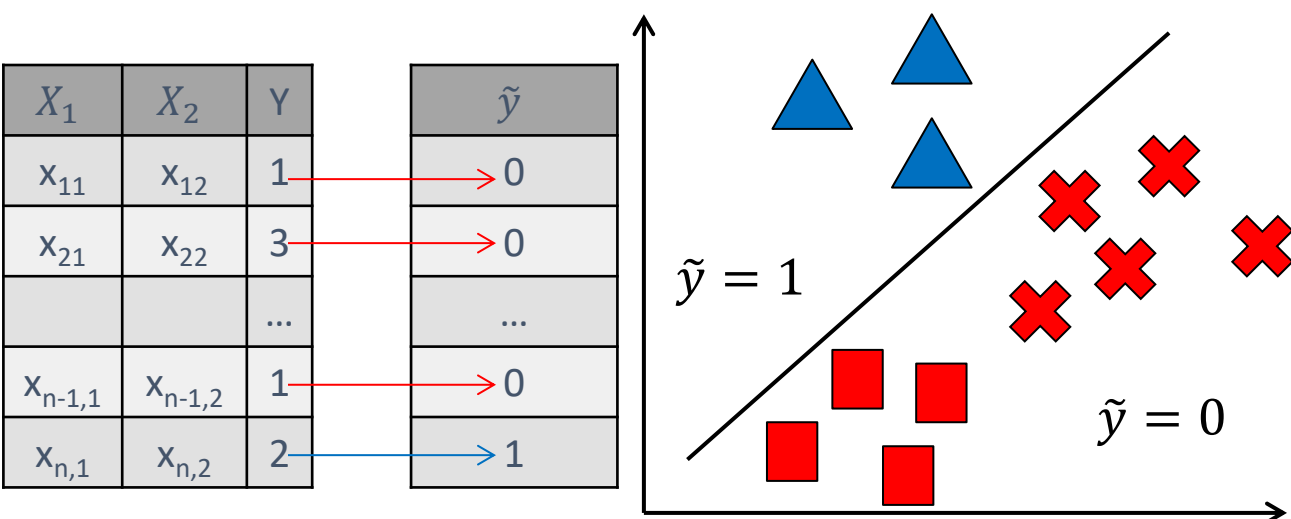
$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (\tilde{y}_i \ln(h_{\theta}(x_{ij})) + (1 - \tilde{y}_i) \ln(1 - h_{\theta}(x_{ij}))) + \frac{\lambda}{2n} \sum_{j=1}^k \theta_j^2 \rightarrow \min$$

$$\theta_{[3 \times k]}^1, h_{\theta}^1(X)$$

Обучение с учителем: Логистическая регрессия

Многоклассовая классификация на примере метода «один против всех» (one-vs-all)

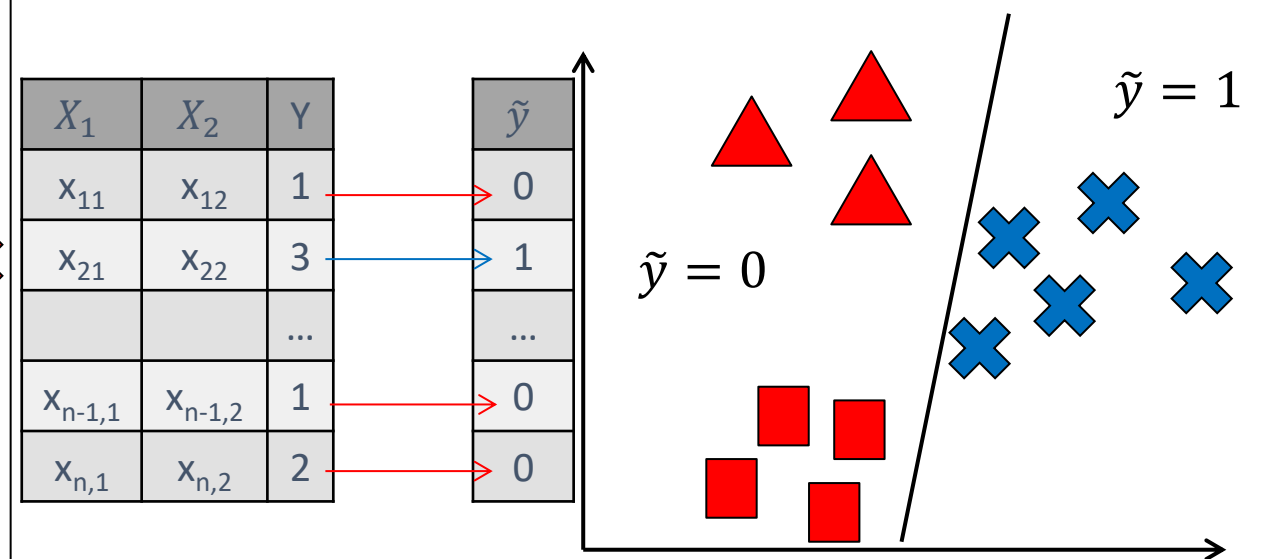
2 шаг «2 класс против всех»



$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (\tilde{y}_i \ln(h_{\theta}(x_{ij})) + (1 - \tilde{y}_i) \ln(1 - h_{\theta}(x_{ij}))) + \frac{\lambda}{2n} \sum_{j=1}^k \theta_j^2 \rightarrow \min$$

$$\theta_{[3 \times k]}^2, h_{\theta}^2(X)$$

3 шаг «3 класс против всех»



$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n (\tilde{y}_i \ln(h_{\theta}(x_{ij})) + (1 - \tilde{y}_i) \ln(1 - h_{\theta}(x_{ij}))) + \frac{\lambda}{2n} \sum_{j=1}^k \theta_j^2 \rightarrow \min$$

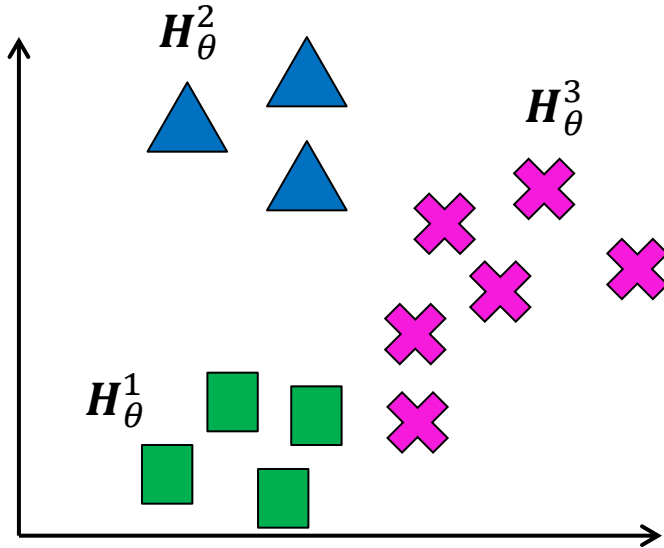
$$\theta_{[3 \times k]}^3, h_{\theta}^3(X)$$

Обучение с учителем: Логистическая регрессия

Многоклассовая классификация на примере метода «один против всех» (one-vs-all)

$X \in R, Y \in \{1; 2; 3\}$

X	Y
x_1	1
x_2	3
...	...
x_{n-1}	1
x_n	2



Определение класса для объекта с максимальным значением $h^q(X)$:

$$h_{\theta}^q(x) = \frac{1}{1 + \exp(-\sum_{i=0}^m \theta_i^q x_i)}$$

	1 класс	2 класс	3 класс	
X	$h^1(X)$	$h^2(X)$	$h^3(X)$	
x_1	0,3	0,7	0,2	$\max_q(h_{\theta}^q(x_1))$
x_2				
...				
x_{n-1}				
x_n				$\max_q(h_{\theta}^q(x_n))$