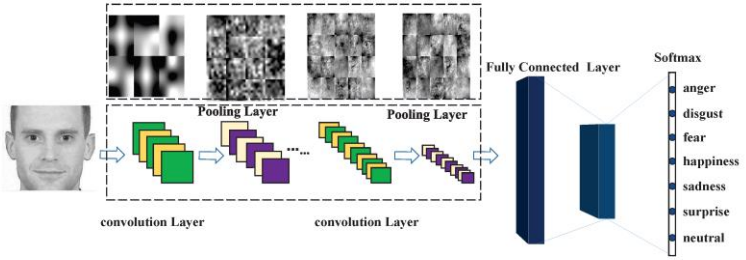
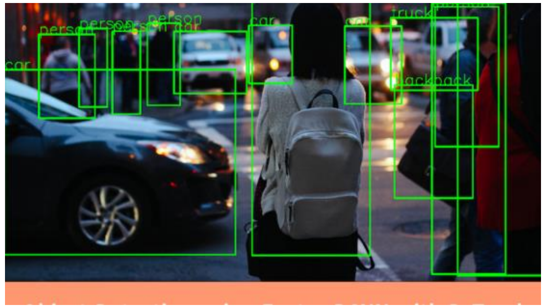
Основы искусственного интеллекта





Лекция

Обучение с учителем: Логистическая регрессия

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

Обучение с учителем: Логистическая регрессия

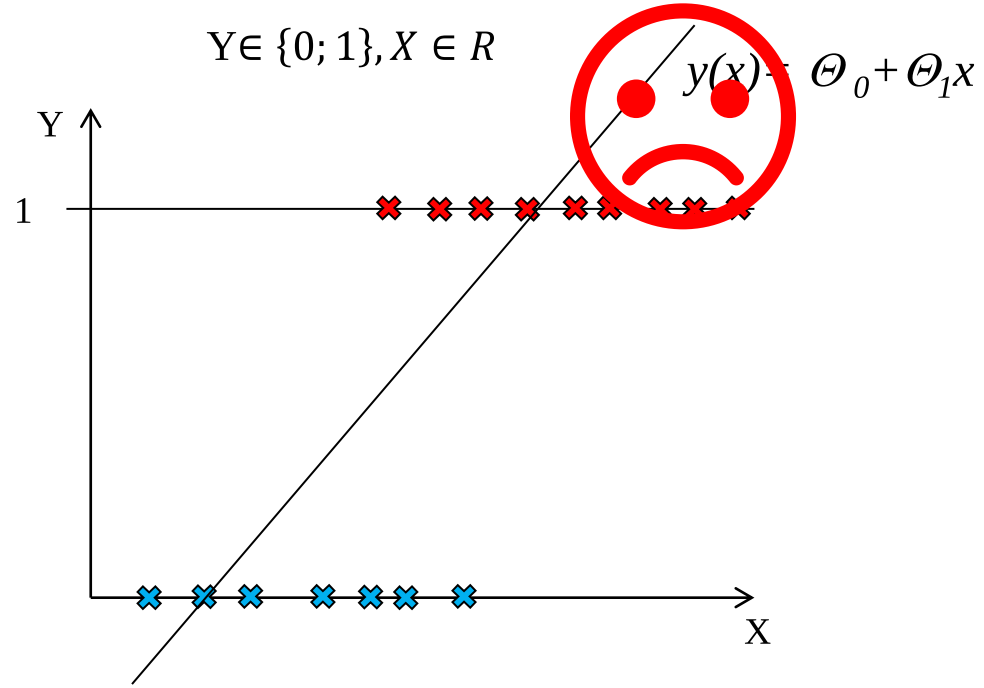
Постановка задачи классификации

Имеются k классов, условно обозначенных y=0, y=1,.. 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

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

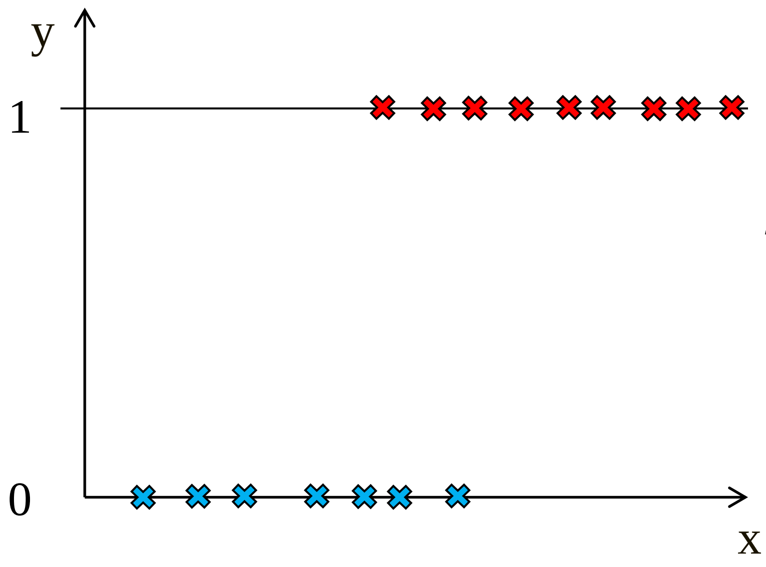


Обучение с учителем: Логистическая регрессия

Постановка задачи классической регрессии

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

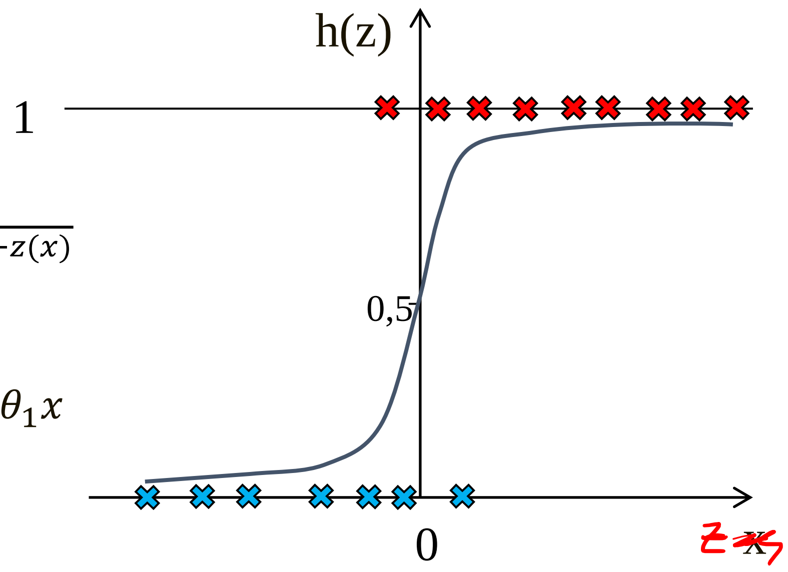
$Y \in \{0; 1\}, X \in R$



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

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

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



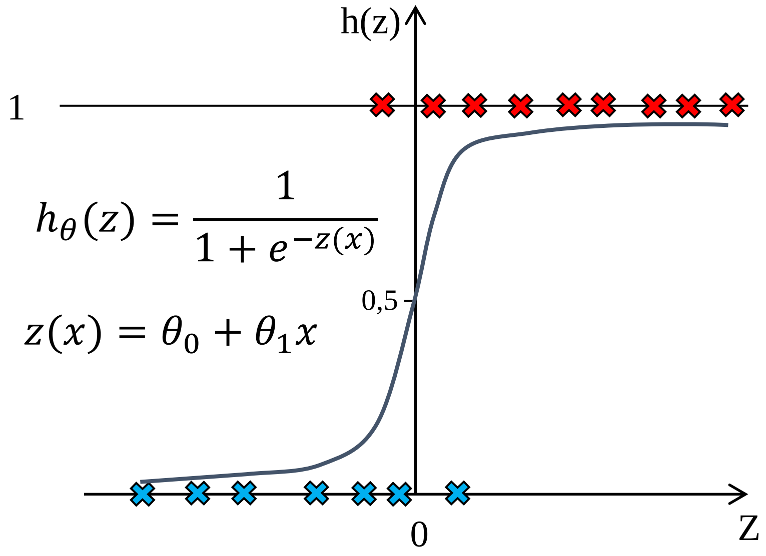
Обучение с учителем: Логистическая регрессия

Постановка задачи классической регрессии

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

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

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



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

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

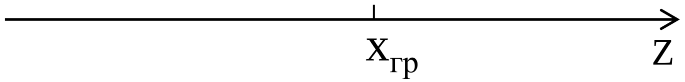
$\mathrm{h(z) = P(y = 1)} \in [0; 1]$ – вероятность принадлежности к классу «1»:

- 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.

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

или

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



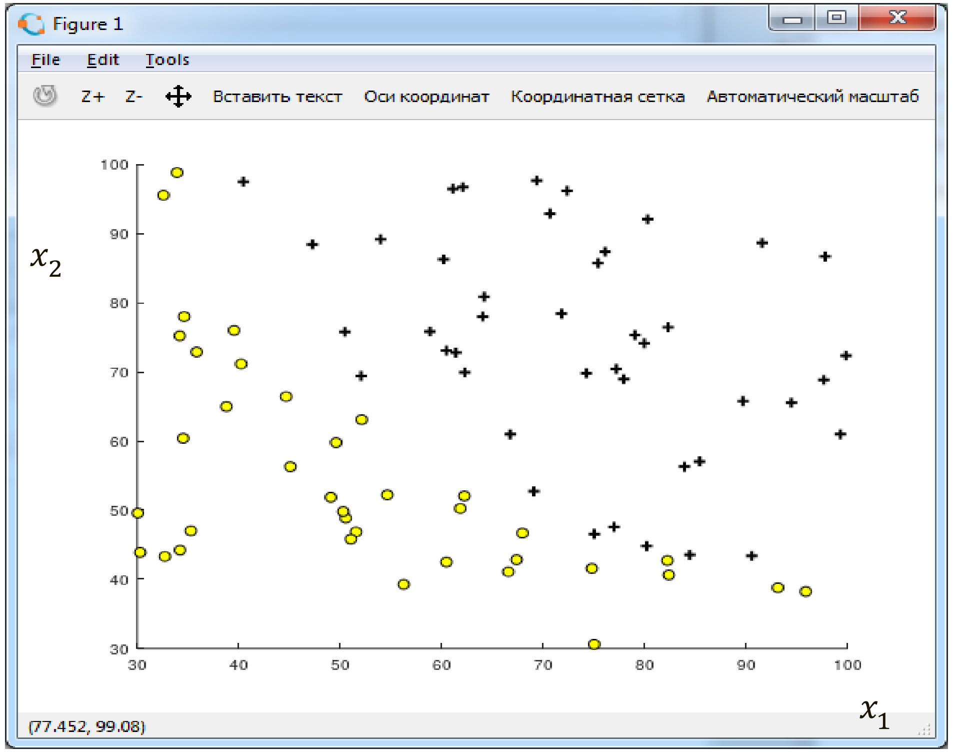
$$z(x) = 0\colon\ \theta_0 + \theta_1 x_{гр} = 0 => x_{гр} = -\frac{\theta_0}{\theta_1}$$

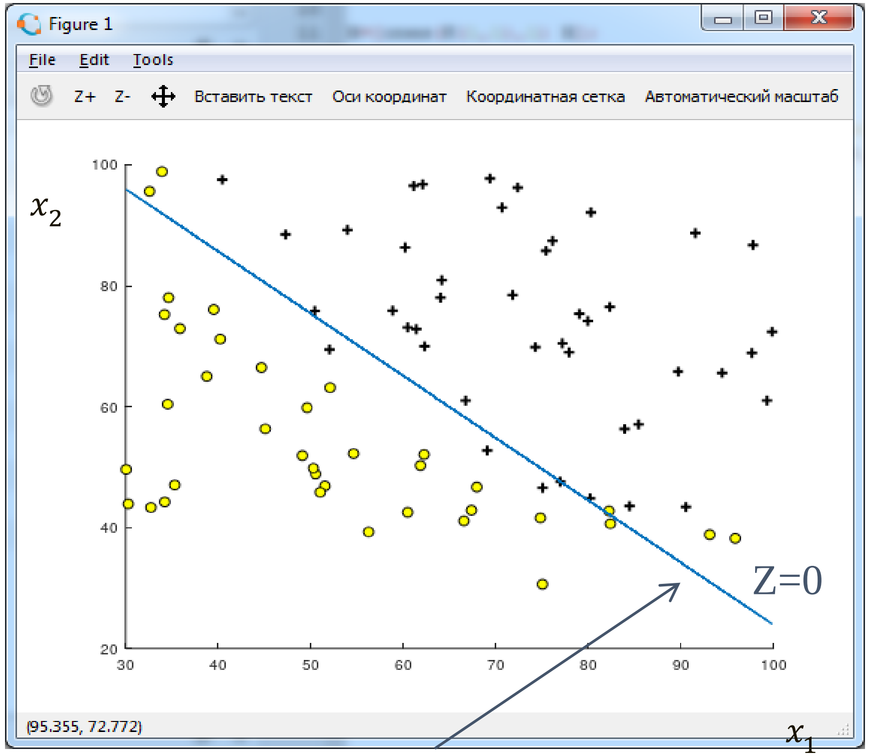
Обучение с учителем: Логистическая регрессия

Постановка задачи классической регрессии

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

<u>Пример</u> Двумерный случай: $\boldsymbol{X} = [[X_1, X_2]]$





$$z(\boldsymbol{X}) = 0,$$

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

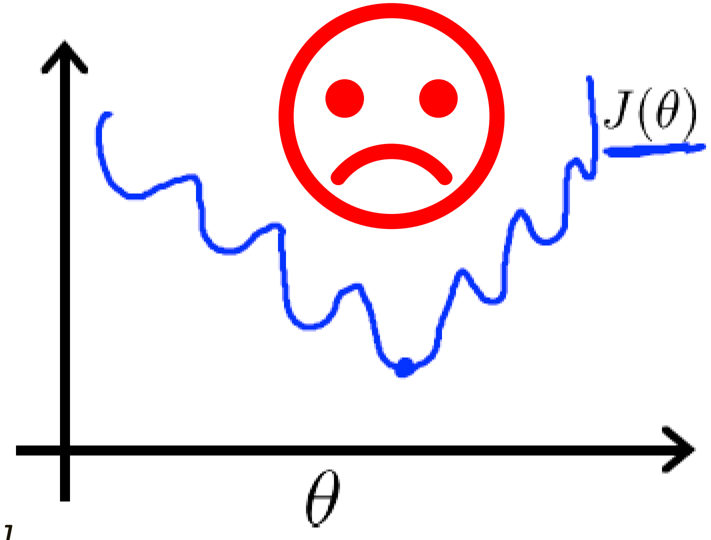
Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели

$$h_\theta(x) = \frac{1}{1+\exp(-z(x))},\ \boldsymbol{\Theta} = [\![\Theta_i]\!]$$ - параметры модели

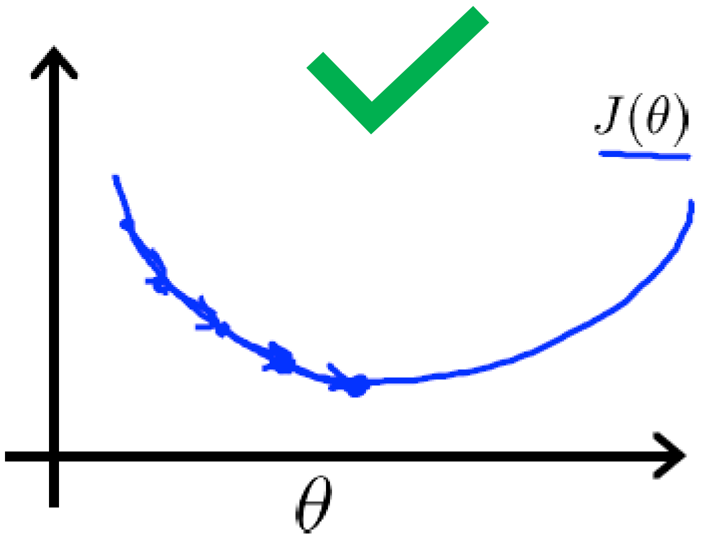
Задача безусловной оптимизации функции ошибки:

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



[Andrew Ng]

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



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.

sum(y_i==1) = n_1

p_i - вероятность события y_i=1

sum(y_i==0) = n_2

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

$$P(y_i) = p_i^{y_i}(1 - p_i)^{1-y_i} = \left[\begin{array}{l} p_i \text{, если } y_i = 1 \\ 1 - p_i \text{, если } y_i = 0 \end{array}\right.$$

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели

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

В n независимых испытаниях, в которых событие y_i=1 может появиться либо нет.
Вероятность p наступления события постоянна.

sum(y_i==1) = n_1	sum(y_i==0) = n_2
$p_i = h(\theta, y_i)$- вероятность события y_i=1	$q_i = 1 - p_i$ - вероятность события y_i=0

Функция максимального правдоподобия:

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

Функция ошибки:

$$J(\theta) = -\ln(L(\boldsymbol{\theta}, \boldsymbol{Y})) \underset{\theta}{\rightarrow} min$$

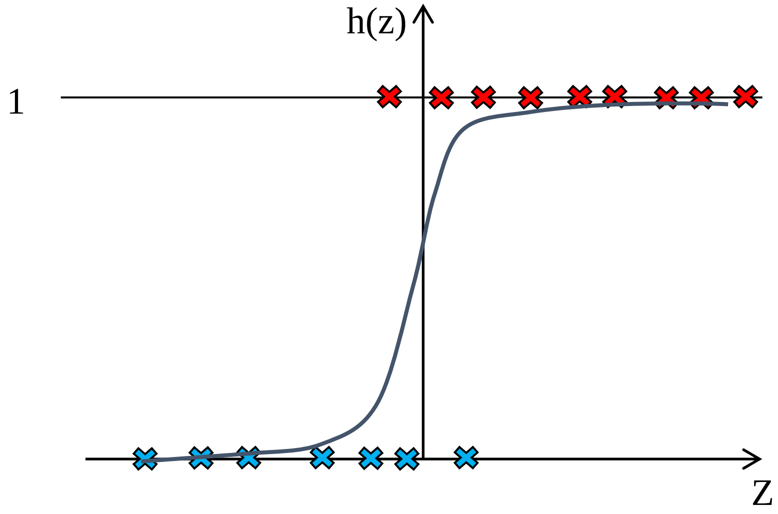
$$\boxed{J(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}(y_i \ln(h_i) + (1-y_i)(\ln(1-h_i)) \rightarrow \min}$$

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели

Функция ошибки:

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



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

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

<u>Примечание</u>: в качестве $h_\theta(x)$ могут быть другие функции из семейства сигмоидных

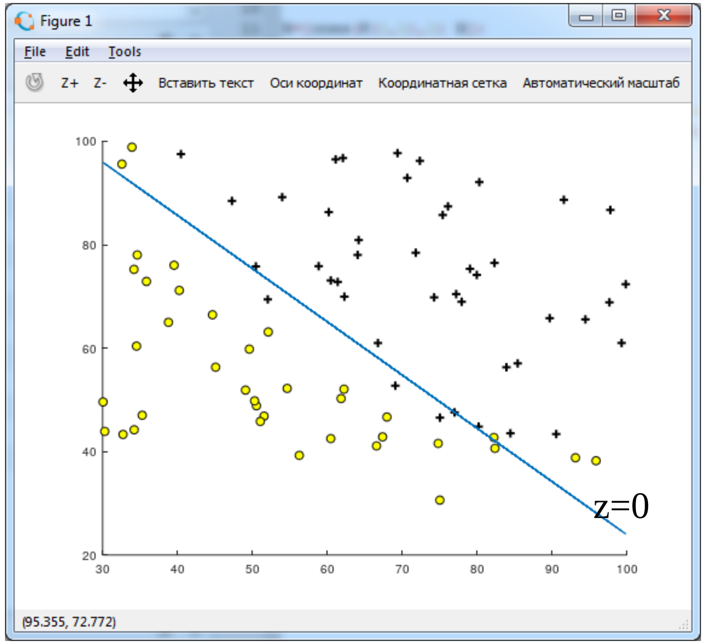
Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

I. Линейная граница между классами

Пример Линейная граница для 2-d случая

$$z(\boldsymbol{X}) = \theta_0 + \theta_1 \boldsymbol{X}_1 + \theta_2 \boldsymbol{X}_2$$



Общий случай Линейная граница для многомерного случая:

$$z(\boldsymbol{X}) = \theta_0 + \theta_1 \boldsymbol{X}_1 + .. + \theta_m \boldsymbol{X}_m = \theta_i \boldsymbol{X}_i\ ^*,$$

где $\boldsymbol{X} = [\![X_i]\!]$ - факторное пространство;
$\boldsymbol{X}_0 = [[1 \ldots 1]]^T$.

Логистическая регрессия:

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

*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:

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

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

Необходимое условие экстремума:

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

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

$\boldsymbol{X}$

$\boldsymbol{X}_{[n\times(m+1)]},\quad \boldsymbol{Y}_{[n\times1]},\quad \boldsymbol{\Theta}_{[m+1,1]},$

$\boldsymbol{Z}_{[n\times1]} = \boldsymbol{Z}(\boldsymbol{X}),\quad \boldsymbol{H}_{[n\times1]} = sigm(\boldsymbol{Z})$

В покомпонентном виде:

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

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

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

В матричном виде:

$$\boxed{\underbrace{\boldsymbol{Z}}_{[n\times1]} = \boldsymbol{X\Theta},}$$

$$\boxed{\underbrace{\boldsymbol{H}}_{[n\times1]} = sigm(\boldsymbol{Z})_{[n\times1]}}$$

$$\boxed{J(\boldsymbol{\theta}) = -\frac{1}{n}(\boldsymbol{Y}^T \ln \boldsymbol{H} + (\boldsymbol{I}-\boldsymbol{Y})^T \ln(\boldsymbol{I}-\boldsymbol{H})) \underset{\boldsymbol{\Theta}}{\rightarrow} min}$$

где $\boldsymbol{I} = (1)_{[n\times1]}$

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

$$\boldsymbol{X}_{[n\times(m+1)]},\quad \boldsymbol{Y}_{[n\times1]},\quad \boldsymbol{\Theta}_{[m+1,1]},$$

$$\boldsymbol{Z}_{[n\times1]} = \boldsymbol{Z}(\boldsymbol{X}),\quad \boldsymbol{H}_{[n\times1]} = sigm(\boldsymbol{Z})$$

Аналитическое решение:

$$\boxed{\boldsymbol{Z} = \boldsymbol{X\Theta},}$$

$$\boxed{J = -\frac{1}{n}(\boldsymbol{Y}^T \ln \boldsymbol{H} + (\boldsymbol{I} - \boldsymbol{Y})^T \ln(\boldsymbol{I} - \boldsymbol{H})) \underset{\boldsymbol{\Theta}}{\to} min}$$

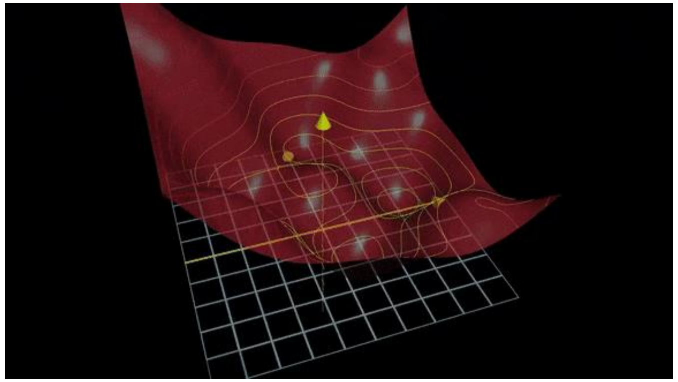
для относительно небольшого кол-ва признаков и объема выборки можно использовать аналитическое решение:

$$\boxed{\boldsymbol{\Theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}}$$

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

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



Gradient descent intuition

Компоненты градиента:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{n}\sum_{i=1}^{n}\left(h_\theta(x_{ij}) - y_i\right) x_{ij}$$

$$\nabla \boldsymbol{J} = \frac{1}{n}\boldsymbol{X}^T\,(\boldsymbol{H} - \boldsymbol{Y}),$$

Расчет коэффициентов на каждом шаге *l+1*:

$$\theta_j^{l+1} = \theta_j^l - \alpha\frac{\partial J}{\partial \theta_j}|_l$$

$$\boldsymbol{\Theta}^{l+1} = \boldsymbol{\Theta}^l - \alpha\nabla \boldsymbol{J}|_l$$

$\boldsymbol{\Theta}^0 = (\theta_j^0)$ − начальное значение коэффициентов;

$\nabla \boldsymbol{J}|_l = \left(\frac{\partial J}{\partial \theta_j}\right)|_l$ − 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:

$$\nabla J = \frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{n}\sum_{i=1}^{n}\left(h_\theta(x_{ij}) - y_i\right)x_{ij}$$

$$\nabla \boldsymbol{J} = \frac{1}{n}\boldsymbol{X}^T\,(\boldsymbol{H} - \boldsymbol{Y}),$$

Расчет коэффициентов на каждом шаге l:

$$\theta_j^{l+1} = \theta_j^l - \lambda\frac{\partial J}{\partial \theta_j}|_l$$

$$\boldsymbol{\Theta}^{l+1} = \boldsymbol{\Theta}^l - \alpha\nabla\boldsymbol{J}|_l$$

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

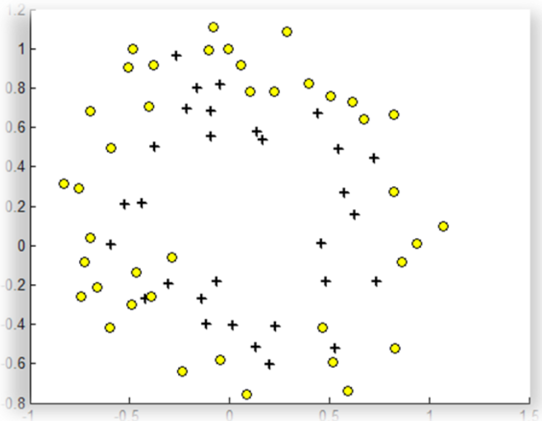
$$J(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left(y_i\ln(h_\theta(x_{ij})) + (1-y_i)(\ln(1-h_\theta(x_{ij})))\right)$$

$$\boxed{J = -\frac{1}{n}(\boldsymbol{Y}^T\ln\boldsymbol{H} + (\boldsymbol{I}-\boldsymbol{Y})^T\ln(\boldsymbol{I}-\boldsymbol{H}))}$$

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

II. Нелинейная граница между классами. Линеаризация



$z(\boldsymbol{X})$– полином степени d.

Пример для случая d=2, m=2.

$z(\boldsymbol{X}) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_1 X_2 + \theta_4 X_1^2 + \theta_5 X_2^2$ пусть $X_3=X_1X_2$, $X_4=(X_1)^2$, $X_5=(X_2)^2$

	X_0	X_1	X_2	X_3	X_4	X_5	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}	$x_{n1}x_{n2}$	$(x_{21})^2$	$(x_{22})^2$	y_n

$X[n \times k]$

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

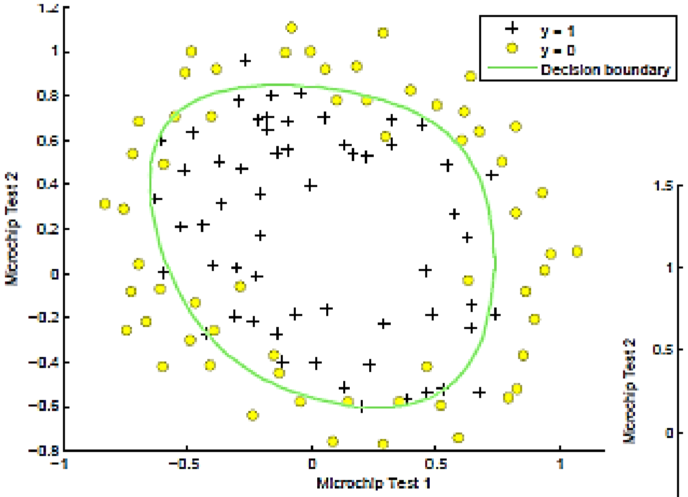
$$\boldsymbol{\theta} = (\theta_i)_{k\times 1}$$

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

Обучение с учителем: Логистическая регрессия

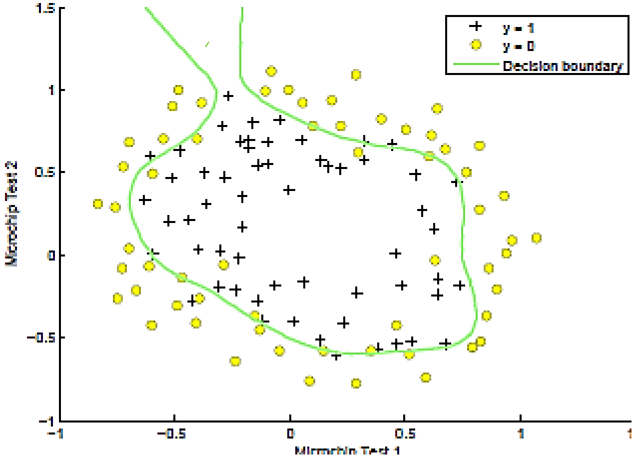
Логистическая регрессия. Определение параметров модели на примере сигмоиды

II. Нелинейная граница между классами. Регуляризация



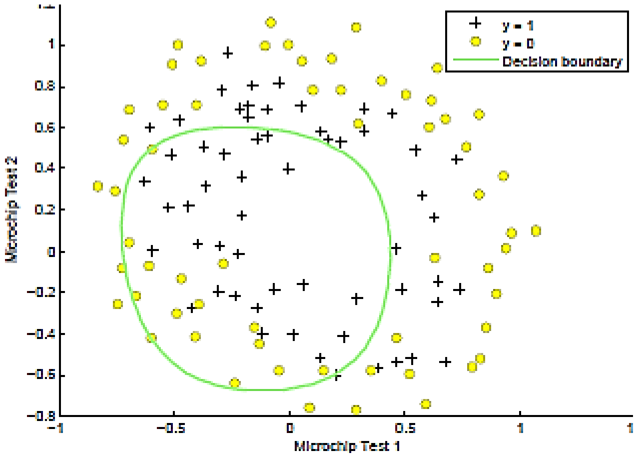
$\lambda = 1$

Переобучение



$\lambda = 0$

$\lambda = 100$



Недообучение

L_2 – регуляризация:

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

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

II. Нелинейная граница между классами. Регуляризация

Функция ошибки:

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

Компоненты градиента функции ошибки:

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

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

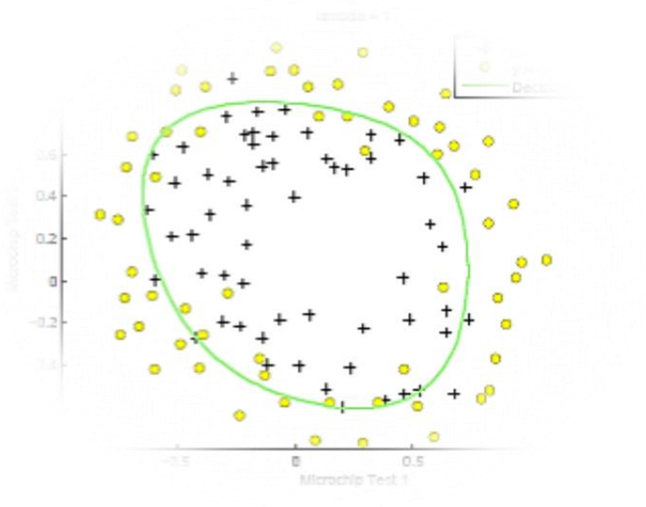
$$\nabla \boldsymbol{J} = \frac{1}{n}\boldsymbol{X}^T\ (\boldsymbol{h} - \boldsymbol{Y}) + \frac{\lambda}{n}\boldsymbol{\theta}$$

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

Обучение с учителем: Логистическая регрессия

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:

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

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

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

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

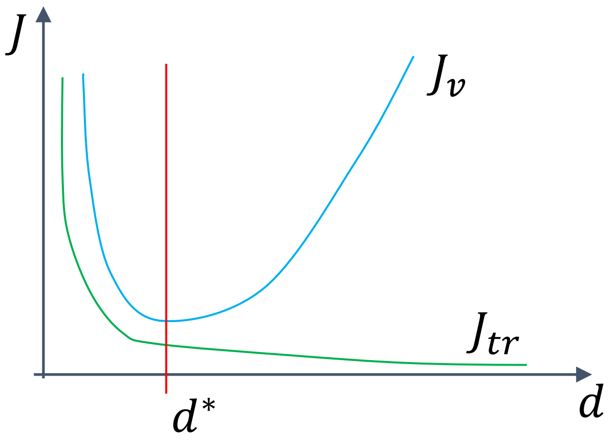
- по значениям J_{tr} и J_v выбирается (d_i, λ_j).

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

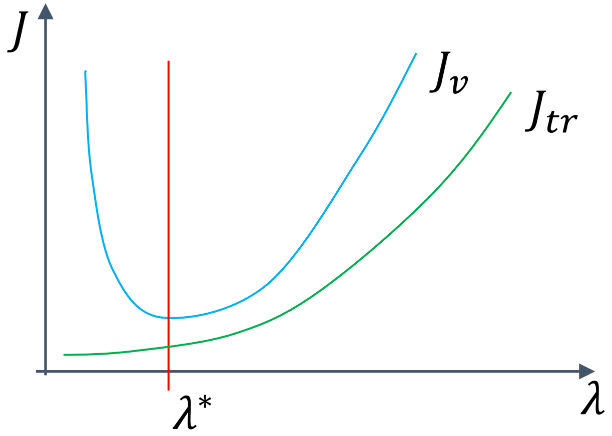
III. Регуляризация. Подбор гиперпараметров модели

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



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

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



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

Обучение с учителем: Логистическая регрессия

Логистическая регрессия. Определение параметров модели на примере сигмоиды

III. Регуляризация. Подбор гиперпараметров модели

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

- Выборка делится на три части: для обучения, валидации и тестирования;
- По обучающей выборке параметры модели $\boldsymbol{\theta} = (\theta_i)_{k\times 1}$ 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 $\left(d_i, \lambda_j\right)$
- Рассчитывается точность модели (см. далее) на тестовой выборке.

Обучение с учителем: Логистическая регрессия

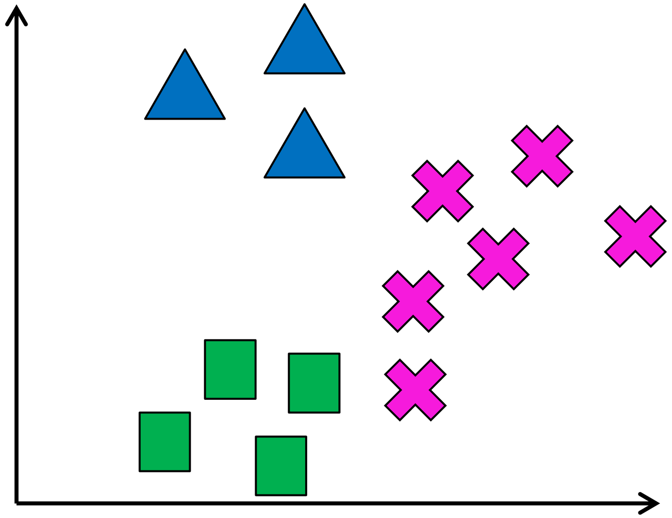
Многоклассовая классификация на примере метода «один против всех» (one-vs-all)

Метод «Один против всех»

Пример для трех классов.

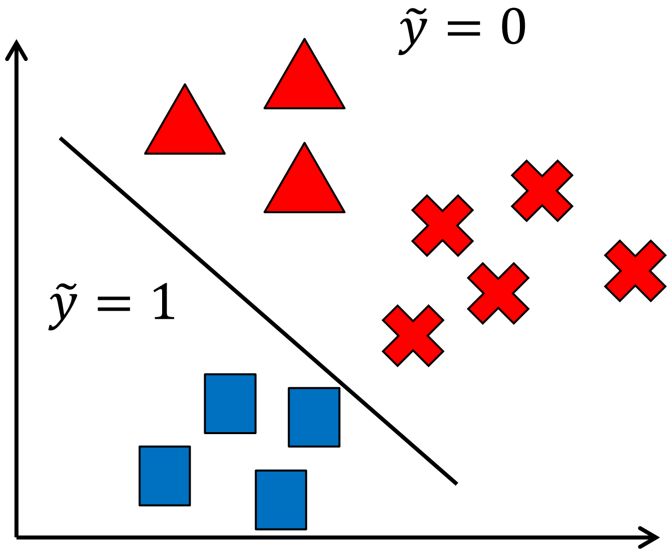
$$X \in R, Y \in \{1; 2; 3\}$$

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



I шаг «1 класс против всех»

X_1	X_2	Y	$\tilde{y}$
x_{11}	x_{12}	1	1
x_{21}	x_{22}	3	0
		…	…
$x_{n-1,1}$	$x_{n-1,2}$	1	1
$x_{n,1}$	$x_{n,2}$	2	0



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

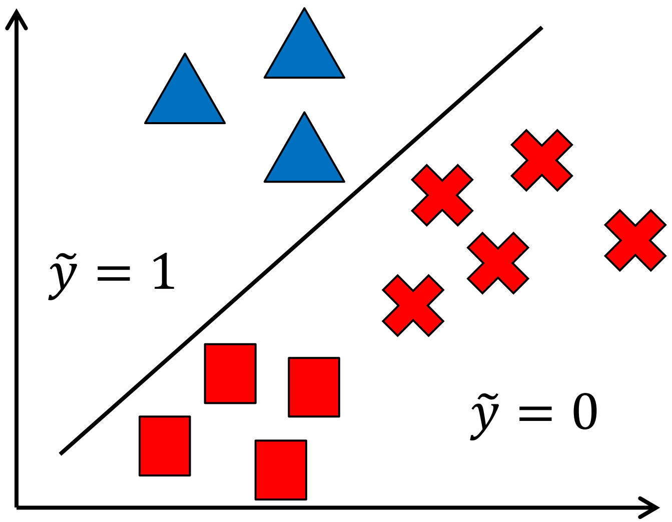
$$\theta^1_{[3\times k]}, h^1_\theta(X)$$

Обучение с учителем: Логистическая регрессия

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

<u>2 шаг «2 класс против всех»</u>

X_1	X_2	Y	$\tilde{y}$
x_{11}	x_{12}	1	0
x_{21}	x_{22}	3	0
		…	…
$x_{n-1,1}$	$x_{n-1,2}$	1	0
$x_{n,1}$	$x_{n,2}$	2	1

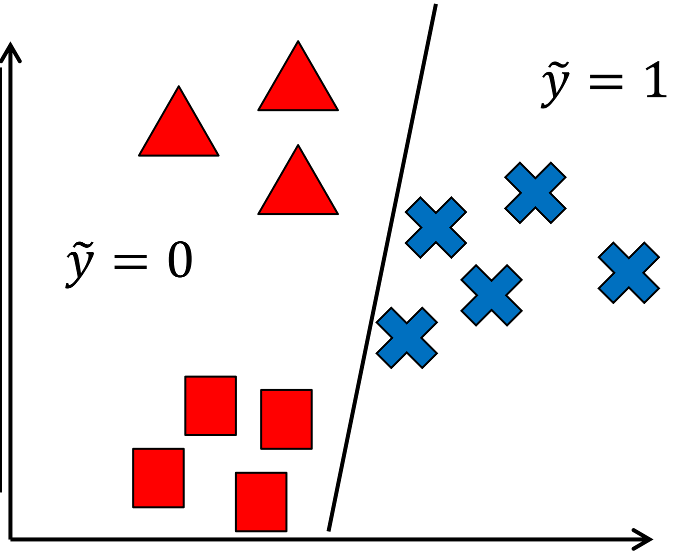


$$J(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{y}_i \ln(h_\theta(x_{ij})) + (1-\tilde{y}_i)(\ln(1-h_\theta(x_{ij})))\right) + \frac{\lambda}{2n}\sum_{j=1}^{k}\theta_j^2 \to \min$$

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

<u>3 шаг «3 класс против всех»</u>

X_1	X_2	Y	$\tilde{y}$
x_{11}	x_{12}	1	0
x_{21}	x_{22}	3	1
		…	…
$x_{n-1,1}$	$x_{n-1,2}$	1	0
$x_{n,1}$	$x_{n,2}$	2	0



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

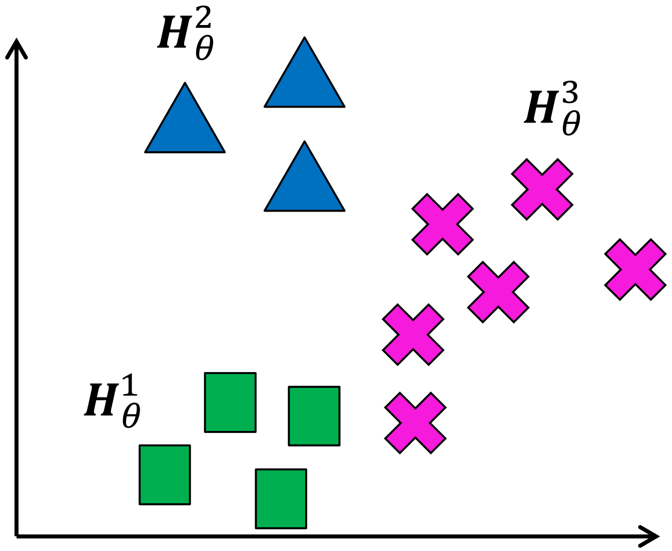
$$\theta^3_{[3\times k]}, h^3_\theta(X)$$

Обучение с учителем: Логистическая регрессия

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

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

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			
x_2			
x_{n-1}			
x_n			

$\max\limits_q(h_\theta^q(x_1))$

…

$\max\limits_q(h_\theta^q(x_n))$