Машинное обучение (часть 1)
Лекция 4. Логистическая регрессия

Заливако Сергей Сергеевич

Кафедра Информатики (402-5)
zalivako@bsuir.by

Примеры задач

- Определение спама в электронных письмах;
- Определение, является ли транзакция мошеннической;
- Определение злокачественности опухоли.

Бинарная классификация $y \in \{0, 1\}$

- 0 – "отрицательный" класс (например, опухоль является доброкачественной);
- 1 – "положительный" класс (например, опухоль является злокачественной);

Многоклассовая классификация $y \in \{0, 1, 2, 3\}$

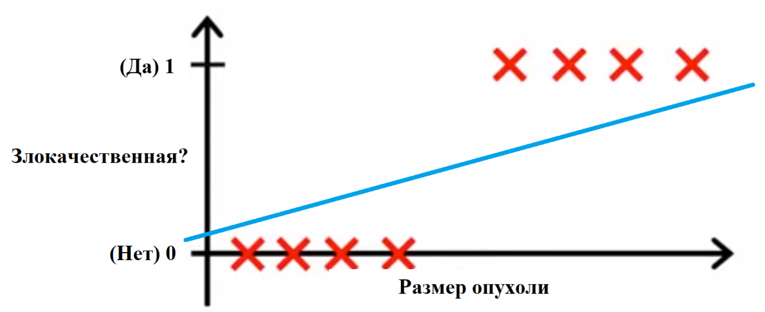
- 0 – объект на картинке является пешеходом;
- 1 – объект на картинке является дорожным знаком;
- 2 – объект на картинке является автомобилем;
- 3 – объект на картинке не опознан;

Задача определение злокачественности опухоли

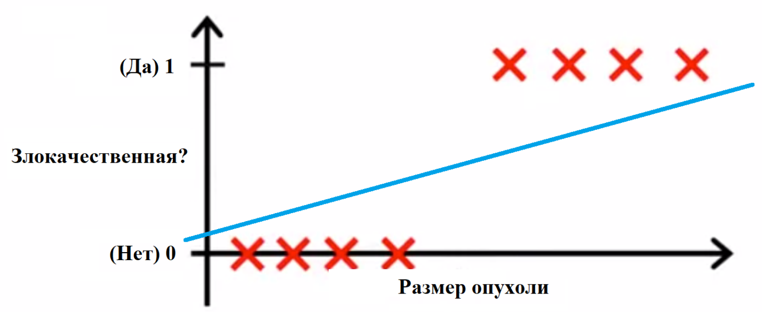
Задача определение злокачественности опухоли

Как решить с помощью линейной регрессии?

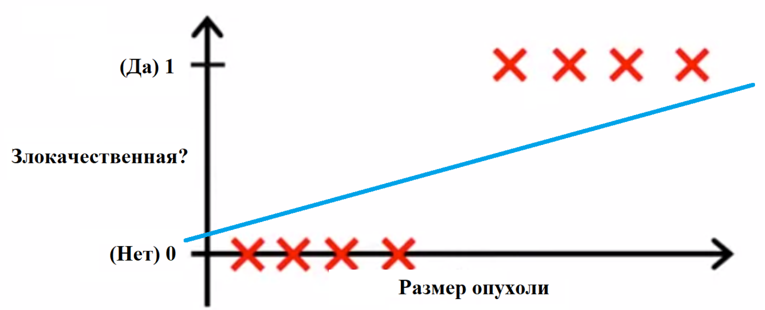
Решение с помощью линейной регрессии



Решение с помощью линейной регрессии

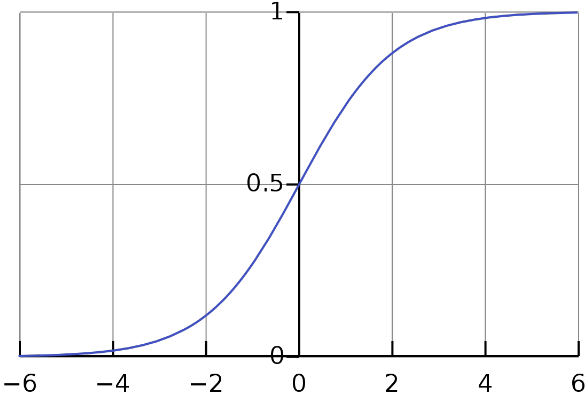


Если $h_\theta(x) \geq 0.5$, то $y = 1$,
Если $h_\theta(x) < 0.5$, то $y = 0$.

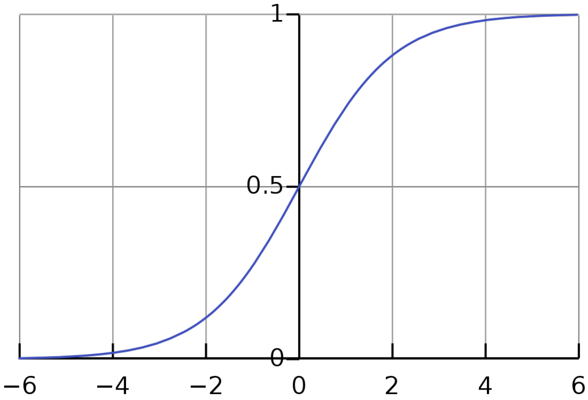


Если $h_\theta(x) \geq 0.5$, то $y = 1$,
Если $h_\theta(x) < 0.5$, то $y = 0$.
Проблема в том, что $h_\theta(x)$ может принимать значения <0 и >1.

Логистическая функция (сигмоид)



Логистическая функция (сигмоид)



$f(z) = \frac{1}{1+e^{-z}}$

Интерпретация логистической функции

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

$y = 1$, если $h_\theta(x) \geq 0.5$

Интерпретация логистической функции

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

$y = 1$, если $h_\theta(x) \geq 0.5$
Следовательно $\theta^T x \geq 0$.

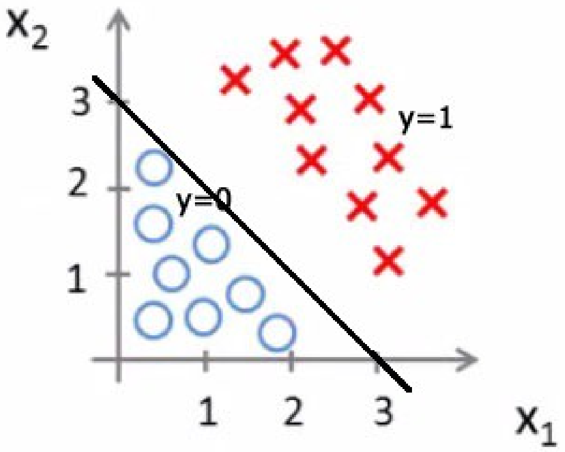
$y = 0$, если $h_\theta(x) < 0.5$
Следовательно $\theta^T x < 0$.

Гипотеза для логистической регрессии

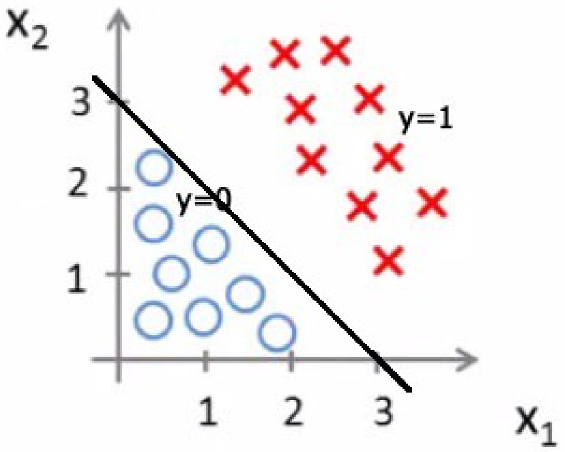
Значение $0 \le h_\theta(x) \le 1$.
Таким образом, $h_\theta(x)$ оценивает **вероятность** того, что объект x принадлежит "положительному" классу $(y = 1)$.

Другими словами, $h_\theta(x) = P(y = 1|x; \theta)$;
$P(y = 0|x; \theta) + P(y = 1|x; \theta) = 1$

Разделяющая поверхность

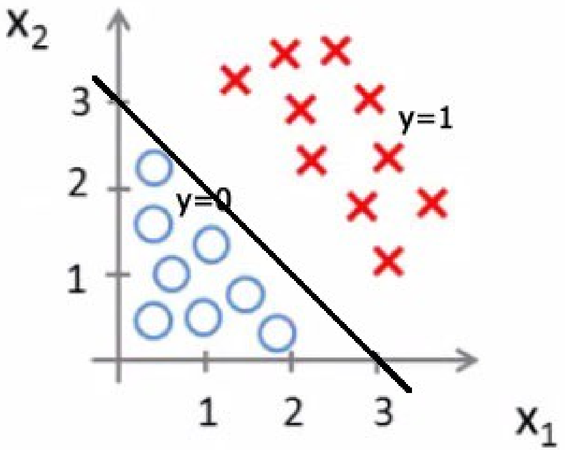


Разделяющая поверхность



Уравнение данной поверхности – $\theta^T x$

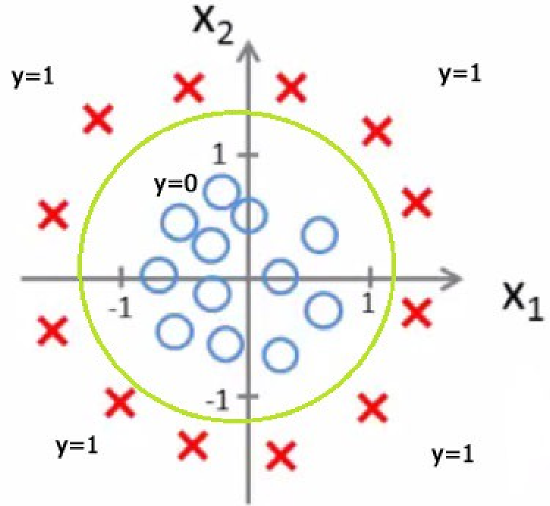
Разделяющая поверхность

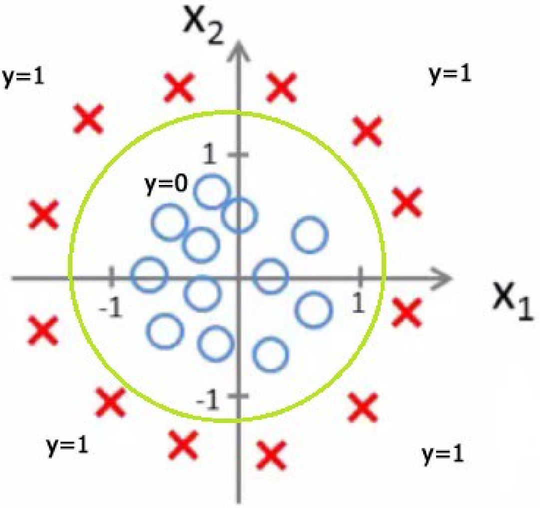


Уравнение данной поверхности – $\theta^T x$

$y = 1$, если $x_1 + x_2 \geq 3$.

Нелинейная разделяющая поверхность





Уравнение данной поверхности – $x_1^2 + x_2^2 \geq 1$.

Модель логистической регрессии

Обучающая выборка

$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})\}$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \ldots \\ x_n \end{bmatrix} \in \mathbb{R}^{n+1}, y \in \{0, 1\}$$

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

Модель логистической регрессии

Обучающая выборка
$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_n \end{bmatrix} \in \mathbb{R}^{n+1}, y \in \{0, 1\}$$

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

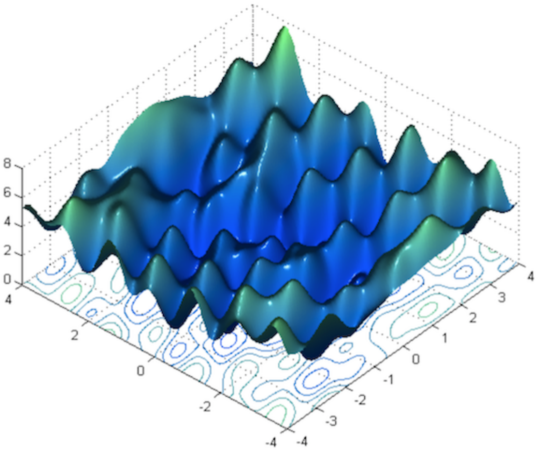
Как выбрать параметры θ?

Квадратичная функция стоимости

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Квадратичная функция стоимости

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

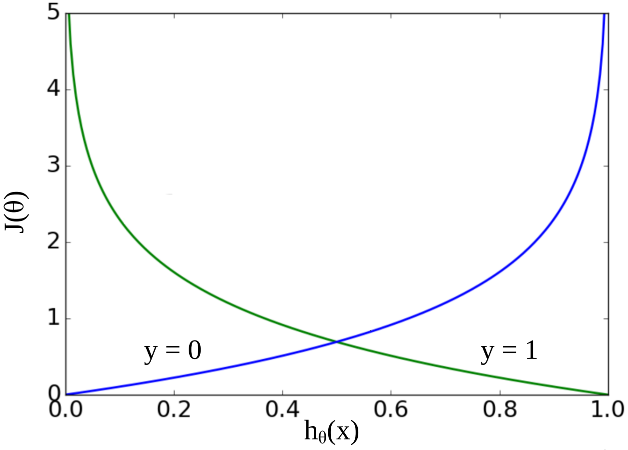


Функция стоимости для логистической регрессии

$Cost(h_\theta(x), y) = -log(h_\theta(x))$, если $y = 1$
$Cost(h_\theta(x), y) = -log(1 - h_\theta(x))$, если $y = 0$

Функция стоимости для логистической регрессии

$$Cost(h_\theta(x), y) = -log(h_\theta(x)), \text{ если } y = 1$$
$$Cost(h_\theta(x), y) = -log(1 - h_\theta(x)), \text{ если } y = 0$$



Объединение двух частей функции потерь

$Cost(h_\theta(x), y) = -log(h_\theta(x))$, если $y = 1$
$Cost(h_\theta(x), y) = -log(1 - h_\theta(x))$, если $y = 0$

Объединение двух частей функции потерь

$Cost(h_\theta(x), y) = -log(h_\theta(x))$, если $y = 1$
$Cost(h_\theta(x), y) = -log(1 - h_\theta(x))$, если $y = 0$

Функция потерь $J(\theta) = \frac{1}{m} \sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$

Объединение двух частей функции потерь

$Cost(h_\theta(x), y) = -log(h_\theta(x))$, если $y = 1$
$Cost(h_\theta(x), y) = -log(1 - h_\theta(x))$, если $y = 0$

Функция потерь $J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$
$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)} log(h_\theta(x^{(i)})) + (1 - y^{(i)}) log(1 - h_\theta(x^{(i)}))]$

$Cost(h_\theta(x), y) = -log(h_\theta(x))$, если $y = 1$
$Cost(h_\theta(x), y) = -log(1 - h_\theta(x))$, если $y = 0$

Функция потерь $J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$
$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}log(h_\theta(x^{(i)})) + (1 - y^{(i)})log(1 - h_\theta(x^{(i)}))]$

Оптимизация $\underset{\theta}{\text{minimize}}\ J(\theta)$

Объединение двух частей функции потерь

$Cost(h_\theta(x), y) = -log(h_\theta(x))$, если $y = 1$
$Cost(h_\theta(x), y) = -log(1 - h_\theta(x))$, если $y = 0$

Функция потерь $J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$
$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}log(h_\theta(x^{(i)})) + (1 - y^{(i)})log(1 - h_\theta(x^{(i)}))]$

Оптимизация $\underset{\theta}{\text{minimize}}\ J(\theta)$

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

Градиентный спуск для логистической регрессии

Пока не достигнута сходимость {

$\theta_j = \theta_j - \alpha\frac{\partial}{\partial\theta_j}J(\theta)$

}

Градиентный спуск для логистической регрессии

Пока не достигнута сходимость {

$\theta_j = \theta_j$ - $\alpha \frac{\partial}{\partial \theta_j} J(\theta)$

}

$j = 0, 1, \ldots, n$

Пока не достигнута сходимость {
$\theta_j = \theta_j - \alpha\frac{\partial}{\partial\theta_j}J(\theta)$
}

$j = 0, 1, \ldots, n$

Формулы для обновления параметров

$\theta_0 = \theta_0 - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}$

$\theta_1 = \theta_1 - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_1^{(i)}$

$\ldots$

$\theta_n = \theta_n - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_n^{(i)}$

Пока не достигнута сходимость {

$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$

}

$j = 0, 1, \ldots, n$

Формулы для обновления параметров

$\theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_0^{(i)}$

$\theta_1 = \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_1^{(i)}$

...

$\theta_n = \theta_n - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_n^{(i)}$

Алгоритм такой же, как и для линейной регрессии!

Применение более сложных методов оптимизации

Пусть дан вектор параметров θ, тогда необходимо определить:

- Процедуру вычисления $J(\theta)$;
- Процедуру вычисления $\frac{\partial}{\partial\theta_j}J(\theta)\forall j \in \{1, 2, \dots, n\}$;

Применение более сложных методов оптимизации

Пусть дан вектор параметров θ, тогда необходимо определить:

- Процедуру вычисления $J(\theta)$;
- Процедуру вычисления $\frac{\partial}{\partial\theta_j}J(\theta)\forall j \in \{1, 2, \ldots, n\}$;

Тогда можно использовать такие методы как:

- BFGS;
- L-BFGS;
- Нелдер-Мид;
- Генетические методы;
- ...

Применение более сложных методов оптимизации

Пусть дан вектор параметров θ, тогда необходимо определить:

- Процедуру вычисления $J(\theta)$;
- Процедуру вычисления $\frac{\partial}{\partial \theta_j} J(\theta) \forall j \in \{1, 2, \ldots, n\}$;

Тогда можно использовать такие методы как:

- BFGS;
- L-BFGS;
- Нелдер-Мид;
- Генетические методы;
- $\ldots$

Преимущества:

- Нет необходимости настраивать α;
- Обычно большая производительность, чем при градиентном спуске.

Недостатки:

- Более сложны в реализации.

Распределение писем по папкам $y \in \{0, 1, 2, 3\}$

- 0 – работа;
- 1 – друзья;
- 2 – семья;
- 3 – хобби.

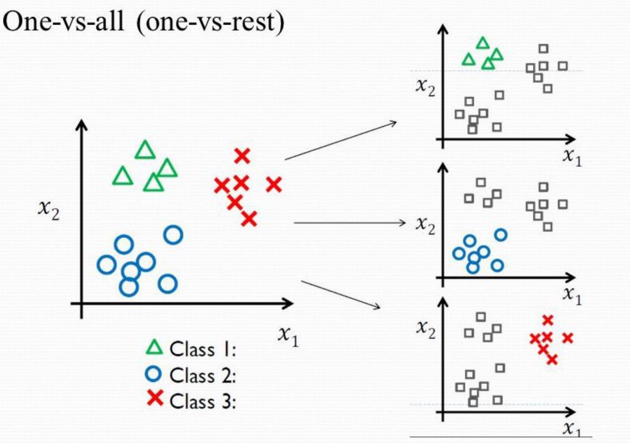
Медицинский диагноз $y \in \{0, 1, 2\}$

- 0 – здоров;
- 1 – простуда;
- 2 – грипп.

Предсказание погоды $y \in \{0, 1, 2, 3\}$

- 0 – солнечно;
- 1 – облачно;
- 2 – дождь;
- 3 – снег;

Метод один против всех



Проблема переобучения

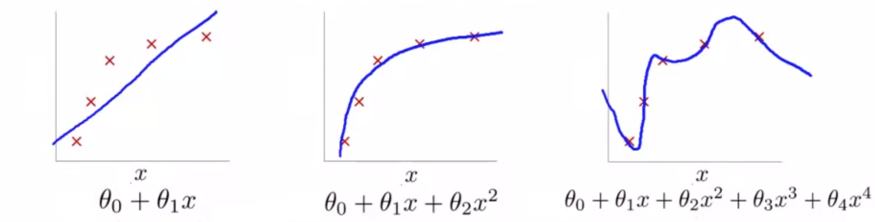
$$\lim_{x \to 8} \frac{1}{x-8} = \infty$$

$$\lim_{x \to 5} \frac{1}{x-5} =$$

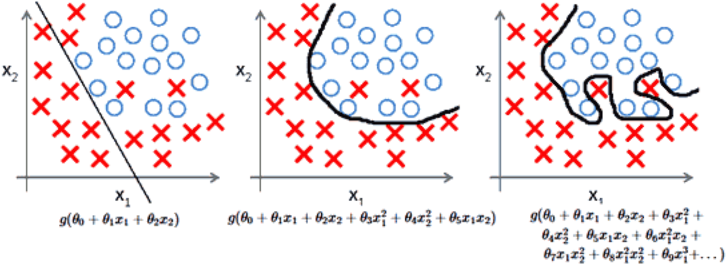
$$\lim_{x \to 8} \frac{1}{x-8} = \infty$$

$$\lim_{x \to 5} \frac{1}{x-5} = 5$$

Переобученная линейная регрессия



Переобученная логистическая регрессия



Функция стоимости

$J(\theta) = \frac{1}{2m}[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^{n} \theta_j^2]$

Регуляризация для линейной регрессии

Функция стоимости

$J(\theta) = \frac{1}{2m}[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^{n} \theta_j^2]$

Градиентный спуск

$\theta_0 = \theta_0$ - $\alpha \frac{1}{m} \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}$

$\theta_1 = \theta_1$ - $\alpha \frac{1}{m} \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_1^{(i)} + \frac{\lambda}{m}\theta_1$

$\ldots$

$\theta_n = \theta_n$ - $\alpha \frac{1}{m} \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_n^{(i)} + \frac{\lambda}{m}\theta_n$

Функция стоимости

$J(\theta) = \frac{1}{2m}[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^{n} \theta_j^2]$

Градиентный спуск

$\theta_0 = \theta_0$ - $\alpha \frac{1}{m} \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}$

$\theta_1 = \theta_1$ - $\alpha \frac{1}{m} \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_1^{(i)} + \frac{\lambda}{m}\theta_1$

$\ldots$

$\theta_n = \theta_n$ - $\alpha \frac{1}{m} \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_n^{(i)} + \frac{\lambda}{m}\theta_n$

Аналитическое решение

$$\theta = (X^T X + \lambda \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix})^{-1} X^T y$$

Функция стоимости $J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m}[y^{(i)} log(h_\theta(x^{(i)})) + (1 - y^{(i)}) log(1 - h_\theta(x^{(i)}))] + \lambda \sum_{j=1}^{n} \theta_j^2]$

Функция стоимости $J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m}[y^{(i)}log(h_\theta(x^{(i)}))+(1-y^{(i)})log(1-h_\theta(x^{(i)}))]+\lambda\sum_{j=1}^{n}\theta_j^2]$

Градиентный спуск

$\theta_0 = \theta_0$ - $\alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}$

$\theta_1 = \theta_1$ - $\alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_1^{(i)} + \frac{\lambda}{m}\theta_j$

$\ldots$

$\theta_n = \theta_n$ - $\alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_n^{(i)} + \frac{\lambda}{m}\theta_j$