Логаритмична регресия

Доц. д-р Ивайло Пенев

Кат. „Компютърни науки и технологии“

Функция на хипотезата

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

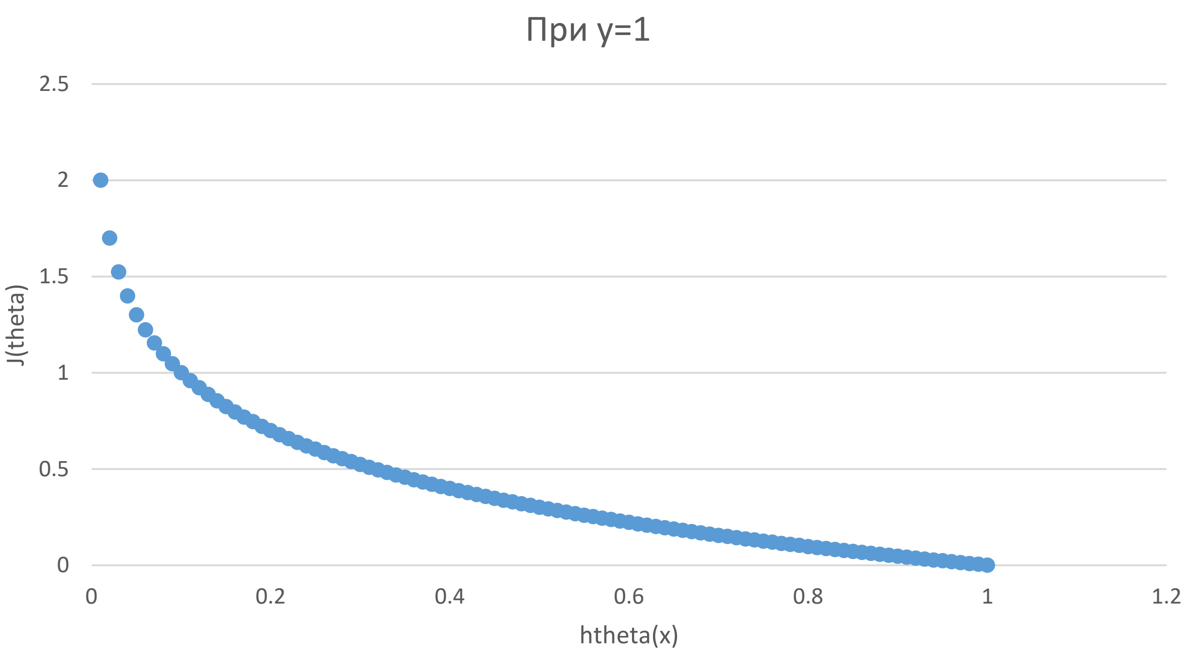
Функция на цената

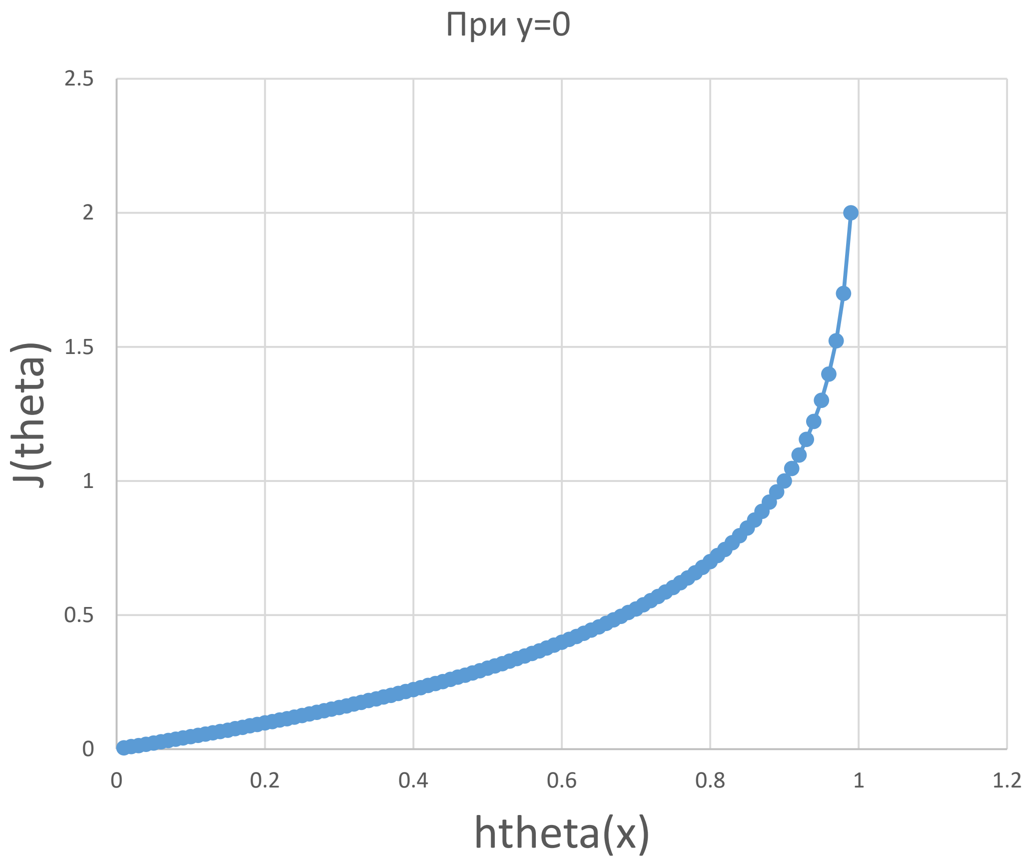
- $J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta\left(x^{(i)}, y^{(i)}\right))$, където

- $Cost(h_\theta\left(x^{(i)}, y^{(i)}\right) = -\log\left(h_\theta(\mathrm{x})\right)$ за $y = 1$

- $Cost(h_\theta\left(x^{(i)}, y^{(i)}\right) = -\log\left(1 - h_\theta(\mathrm{x})\right)$ за $y = 0$

Графично представяне на $J(\theta)$





Обяснение

- $Cost(h_\theta(x), y) = 0$, ако $h_\theta(x) = y$
- $Cost(h_\theta(x), y) \rightarrow \infty$, ако $y = 0$ и $h_\theta(x) \rightarrow 1$
- $Cost(h_\theta(x), y) \rightarrow \infty$, ако $y = 1$ и $h_\theta(x) \rightarrow 0$
- Ако при обучението коректният резултат е $y = 0$, то цената ще бъде 0 при $h_\theta(x)$=0 (т.е. при прогноза 0). Ако прогнозата е по-голяма от 0, то цената ще се увеличава ($Cost(h_\theta(x), y) \rightarrow \infty$)
- Ако при обучението коректният резултат е $y = 1$, то цената ще бъде 0 при $h_\theta(x)$=1 (т.е. при прогноза 1). Ако прогнозата е по-малка от 1, то цената ще се увеличава ($Cost(h_\theta(x), y) \rightarrow \infty$)

Контролен въпрос

В логаритмичната регресия функцията на цената има следният вид:

- $Cost(h_\theta(x), y) = -\log\big(h_\theta(\mathrm{x})\big)$ за $y = 1$
- $Cost(h_\theta(x), y) = -\log\big(1 - h_\theta(\mathrm{x})\big)$ за $y = 0$

Отбележете **всички** верни отговори:

а) Ако $h_\theta(x) = y$, то $cost(h_\theta(x), y) = 0$ за y=0 и y=1

б) Ако y=0, то $cost(h_\theta(x), y) \to \infty$ при $h_\theta(x) \to 1$

в) Ако y=0, то $cost(h_\theta(x), y) \to \infty$ при $h_\theta(x) \to 0$

г) Ако $h_\theta(x) = 0.5$, то $cost(h_\theta(x), y) > 0$ за y=0 и y=1

Опростена функция на цената

- Двата случая на ф-та на цената (за y=0 и y=1) могат да бъдат обединени:
$$Cost(h_\theta(x), y) = -ylog\left(h_\theta(x)\right) - (1-y)\log(1-h_\theta(x))$$
- При y=1 => $Cost(h_\theta(x), y) = -ylog\left(h_\theta(x)\right)$
- При y=0 => $Cost(h_\theta(x), y) = (1-y)\log(1-h_\theta(x))$
- Следователно ф-та на цената може да се представи:
$$\boldsymbol{J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}log(h_\theta\left(x^{(i)}\right) + \left(1-y^{(i)}\right)log(1-h_\theta\left(x^{(i)}\right))]}$$

Векторна форма на функцията на цената

Ако ф-та на цената за всички обучителни примери е:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right)\log(1 - h_\theta\left(x^{(i)}\right))]$$

, то във векторна форма ф-та може да се запише така:

$$h = g(X\theta)$$

$$J(\theta) = \frac{1}{m}.(-y^T\log(h) - (1-y)^T\log(1-h))$$

Градиентно спускане

- Общ вид:

Повтаряй:

{

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

}

- След диференцирането на $J(\theta)$ се получава:

Повтаряй:

{

$$\theta_j = \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta\left(x^{(i)}\right) - y^{(i)}) x_j^{(i)}$$

}

Контролен въпрос

Прилагаме градиентно спускане в логаритмична регресия за намиране на стойностите на параметрите θ. Как ще проверим дали сме избрали подходяща стойност на скоростта на обучение α и дали градиентното спускане работи правилно?

а) Чертаем графиката на ф-та $J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta\left(x^{(i)}\right) - y^{(i)})^2$, като по абсцисната ос нанасяме броя на итерациите. Проверяваме дали стойността на $J(\theta)$ намалява на всяка следваща итерация

б) Чертаем графиката на ф-та $J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta\left(x^{(i)}\right) + \left(1 -\right.$

Контролен въпрос

При градиентно спускане на всяка итерация се изпълняват следните операции:

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

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

.

.

.

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

Искаме да направим векторна реализация на алгоритъма във вида $\theta = \theta - \alpha\beta$, където β е вектор. Изберете правилния вариант на векторната имплементация:

а) $\theta = \theta - \alpha\frac{1}{m}\sum_{i=1}^{m}[(h_\theta(x^{(i)}) - y^{(i)}).x^{(i)}]$

б) $\theta = \theta - \alpha\frac{1}{m}[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)}].x^{(i)}$

в) $\theta = \theta - \alpha\frac{1}{m}x^{(i)}[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)}]$

г) Всички изброени

Векторна форма на градиентното спускане

$$\theta \coloneqq \theta - \frac{\alpha}{m} X^T (g(X\theta) - \vec{y})$$

Класификация за множество класове

- Обобщаваме класификационната задача за случаи, в които имаме повече от две възможности за y, т.е вместо $y = \{0,1\}$ възможните изходи са $y = \{0,1, \dots, n\}$
- Можем да разделим задачата на $n + 1$ бинарни класификационни задачи, във всяка от които предсказваме вероятността y да принадлежи на всеки от n класа, т.е.:

$y = \{0,1, \dots, n\}$

$h_\theta^{(0)}(x) = P(y = 0 \mid x\,;\ \theta)$

$h_\theta^{(1)}(x) = P(y = 1 \mid x\,;\ \theta)$

…

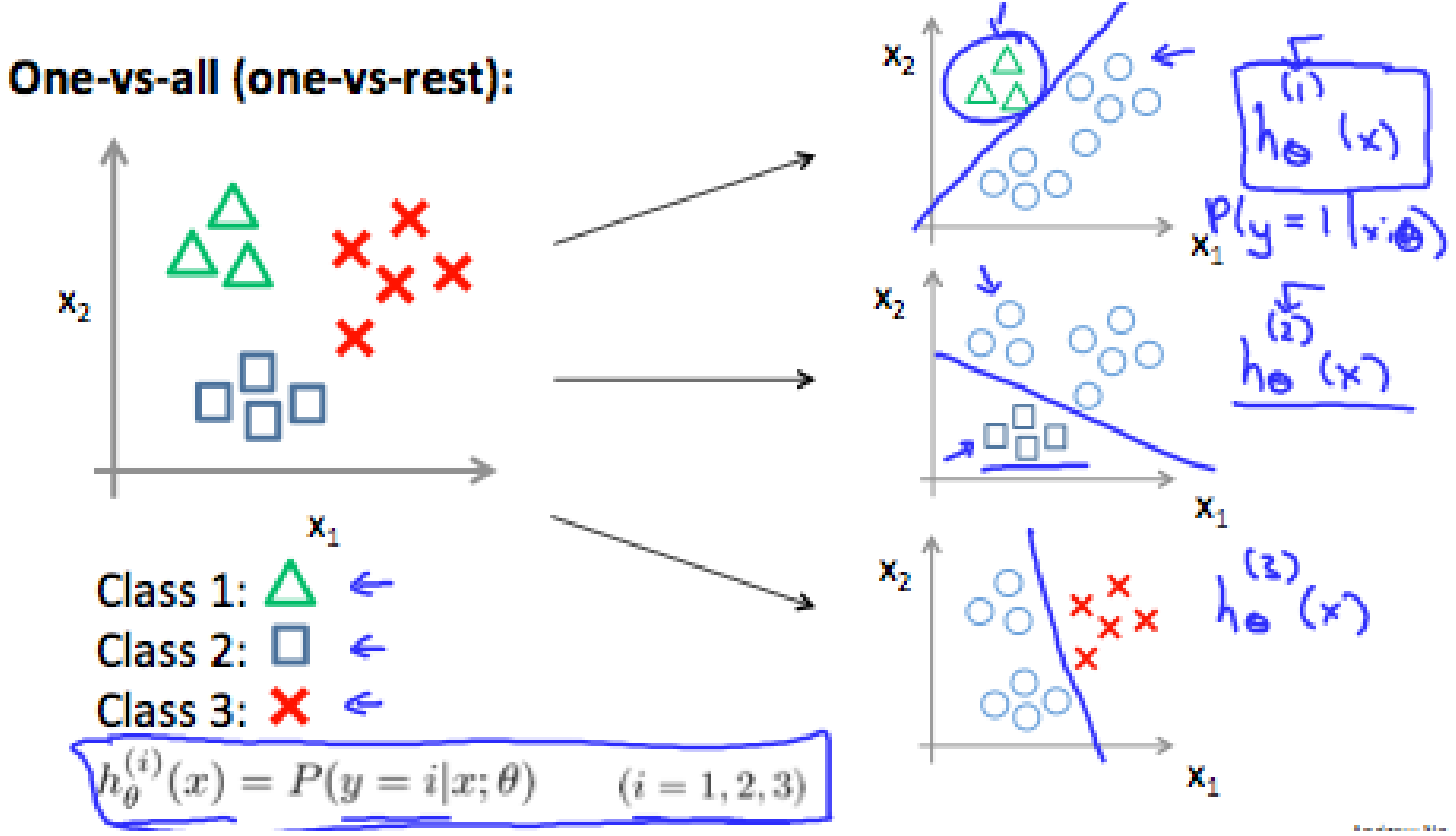
$h_\theta^{(n)}(x) = P(y = n \mid x\,;\ \theta)$

Резултат = $max_i(h_\theta^{(i)}(x))$

Алгоритъм

- Първоначална избираме един клас, а всички останали отделяме в друг клас. Решаваме бинарна класификационна задача. Повтаряме това за всеки възможен клас. Накрая избираме прогнозата с най-висока вероятност.

Пример за класификация с 3 възможни класа



Обобщение

- Обучаваме хипотеза с логаритмична регресия (логаритмичен класификатор) $h_\theta(x)$ за всеки възможен клас *i* – определяме вероятността за *y=i*
- За да определим към кой клас принадлежи неизвестна променлива *x*, избираме класа с максимална $h_\theta(x)$