

Аналитично изчисляване на параметри

Доц. д-р Ивайло Пенев

Кат. „Компютърни науки и технологии“

Нормално уравнение (Normal equation)

- Съществува метод за минимизиране на J без градиентно спускане – „нормално уравнение“
- При този метод J се минимизира чрез директно изчисляване на производните на J спрямо параметрите θ и приравняването им на нула
- Нормалното уравнение има следният вид:

$$\theta = (X^T X)^{-1} X^T y$$

Пример

x_0	Площ x_1	Брой спални x_2	Брой етажи x_3	Възраст на къщата (години) x_4	Цена (\$1000) y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

$$\bullet X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix} - m \times (n+1) \quad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix} - m \times 1$$

$$\bullet \theta = (X^T X)^{-1} X^T y$$

- При нормалното уравнение няма необходимост от мащабиране

Сравнение между градиентно спускане и нормално уравнение

Градиентно спускане	Нормално уравнение
Необходимо е да изберем скорост на обучение alpha	Не е необходимо да избираме alpha
Необходими са много итерации	Не са необходими никакви итерации
Сложност - $O(kn^2)$	Сложност - $O(n^3)$ поради необходимост от изчисляване на обратната матрица $(X^T X)^{-1}$
Работи добре при големи n – голям брой променливи	Работи бавно при големи n

Практически при $n > 10000$ е препоръчително използване на градиентно спускане

Проблем при нормалното уравнение

- **Ако $(X^T X)^{-1}$ не може да се намери**, причините най-често са:
 - Наличие на силно свързани променливи – т.е. променливите имат линейна връзка, напр.:
 - x_1 - площ в $feet^2$
 - x_2 - площ в m^2
 - $1m = 3.28feet$ – линейна връзка, т.е. едната променлива може да се изключи
 - Твърде много променливи (напр. $m \leq n$)
- Решения:
 - Премахване на променлива, която е свързана с друга
 - Намаляване на броя на променливите
 - Регуляризация