Least squares method

Marcin Kuta

Breast cancer dataset

```
1) ID number
2) Diagnosis (M = malignant, B = benign)

3)  radius (mean)
4)  texture (mean)
5)  perimeter (mean)
6)  area (mean)
7)  smoothness (mean)
8)  compactness (mean)
9)  concavity (mean)
10) concave points (mean)
11) symmetry (mean)
12) fractal dimension (mean)
```

Breast cancer dataset

```
13) radius (stderr)
14) texture (stderr)
15) perimeter (stderr)
16) area (stderr)
17) smoothness (stderr)
18) compactness (stderr)
19) concavity (stderr)
20) concave points (stderr)
21) symmetry (stderr)
22) fractal dimension (stderr)
```

Breast cancer dataset

```
23) radius (worst)
24) texture (worst)
25) perimeter (worst)
26) area (worst)
27) smoothness (worst)
28) compactness (worst)
29) concavity (worst)
30) concave points (worst)
31) symmetry (worst)
32) fractal dimension (worst)
```

Least squares method

Linear representation

$$A_{\text{lin}} = \begin{bmatrix} f_{1,1} & \dots & f_{1,m} \\ f_{2,1} & \dots & f_{2,m} \\ \vdots & \dots & \vdots \\ f_{n,1} & \dots & f_{n,m} \end{bmatrix} \tag{1}$$

Least squares method

$$Aw \cong b, \quad A \in \mathbb{R}^{n \times m}, \quad b \in \mathbb{R}^{1 \times n} \tag{2}$$

n - number of instances
m - number of features

$$\min_{w} ||Aw - b|| \tag{3}$$

Least squares method

Linear representation

$$A_{\text{lin}} = \begin{bmatrix} f_{1,1} & f_{1,2} & f_{1,3} & f_{1,4} \\ f_{2,1} & f_{2,2} & f_{2,3} & f_{2,4} \\ \vdots & \vdots & \vdots & \vdots \\ f_{n,1} & f_{n,2} & f_{n,3} & f_{n,4} \end{bmatrix} \tag{4}$$

Quadratic representation

$$A_{\text{quad}} = \begin{bmatrix} f_{1,1}, f_{1,2}, f_{1,3}, f_{1,4}, f_{1,1}^2, f_{1,2}^2, f_{1,3}^2, f_{1,4}^2, f_{1,1}f_{1,2}, f_{1,1}f_{1,3}, f_{1,1}f_{1,4}, f_{1,2}f_{1,3}, f_{1,2}f_{1,4}, f_{1,3}f_{1,4} \\ \vdots \\ f_{n,0}, f_{n,1}, f_{n,2}, f_{n,3}, f_{n,0}^2, f_{n,1}^2, f_{n,3}^2, f_{n,4}^2, f_{n,1}f_{n,2}, f_{n,1}f_{n,3}, f_{n,1}f_{n,4}, f_{n,2}f_{n,3}, f_{n,2}f_{n,4}, f_{n,3}f_{n,4} \end{bmatrix}$$

Normal equation

$$w = \underbrace{(A^T A)^{-1} A^T}_{\substack{\text{pseudoinverse} \\ \text{matrix}}} b \tag{5}$$

$$w = A^{\dagger} b \tag{6}$$

- $A^T A$ is not guaranteed to be non-singular
- The method is overly sensitive to the condition number of matrix
- QR and SVD are numerically more stable alternatives but computationally slower

Normal equation

$$\text{cond}(A) = ||A|| \cdot ||A^{-1}|| \tag{7}$$

Stability of normal equation:

$$\text{cond}(A^T A) = \text{cond}(A)^2 \tag{8}$$

References

[1] Michael T. Heath,
Scientific Computing. An Introductory Survey, 2nd Edition,
Chapter 3: Linear Least Squares
2002

[2] Michael T. Heath,
Chapter 3: Linear Least Squares
http://heath.cs.illinois.edu/scicomp/notes/cs450_chapt03.pdf

[3] Introduction to pandas
https://jakevdp.github.io/PythonDataScienceHandbook/03.00-introduction-to-pandas.html