Chapter 8

Ordinary least squares (OLS). Linear regression.

Ordinary least squares (OLS).

Least squares for implied volatility computation.

Linear regression: ordinary least squares for time series data.

Ordinary least squares for random variables.

The intuition behind ordinary least squares for time series data.

8.1 Ordinary least squares

Let A be an $m \times n$ matrix with more rows than columns, i.e., with $m > n$, and assume that the column vectors of the matrix A are linearly independent. Let y be a column vector of size m.

A solution $x \in \mathbb{R}^n$ to the linear system $Ax = y$ exists if and only if the vector y is a linear combination of the column vectors of A, which is rarely the case in practice.

The ordinary least squares method (OLS) provides an alternative to solving $Ax = y$ exactly, and requires finding a vector $x \in \mathbb{R}^n$ with smallest approximation error $y - Ax$, i.e., such that $||y - Ax||$ is minimal, where $||\cdot||$ denotes the Euclidean norm; see (5.8). This can be stated formally as follows:

$$\text{Given } \; y \in \mathbb{R}^m, \;\; \text{find } \; x \in \mathbb{R}^n \; \text{ such that } \; ||y - Ax|| \; \text{ is minimal.} \tag{8.1}$$

Note that we will also refer to (8.1) as solving the least squares problem

$$y \; \approx \; Ax. \tag{8.2}$$

Problem (8.1) is equivalent to finding the global minimum point of the function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) \; = \; ||y - Ax||^2. \tag{8.3}$$

Recall from (5.8) that $||w||^2 = (w, w) = w^t w$ for any vector w. Then,

$$\begin{aligned} ||y - Ax||^2 &= (y - Ax)^t (y - Ax) \; = \; (y^t - x^t A^t)(y - Ax) \\ &= y^t y - x^t A^t y - y^t A x + x^t A^t A x. \end{aligned} \tag{8.4}$$

Note that

$$y^t Ax = x^t A^t y \tag{8.5}$$

since $(u, v) = (v, u)$ for any u and v and therefore

$$y^t Ax = (Ax, y) = (y, Ax) = (Ax)^t y = x^t A^t y.$$

From (8.4) and (8.5), it follows that

$$||y - Ax||^2 = y^t y - 2x^t A^t y + x^t A^t Ax. \tag{8.6}$$

From (8.3) and (8.6), and since $y^t y = ||y||^2$, we conclude that

$$f(x) = ||y||^2 - 2x^t A^t y + x^t A^t Ax. \tag{8.7}$$

Recall that any minimum point x_0 of $f(x)$ must be a critical point of $f(x)$, i.e., a solution to $Df(x_0) = 0$. From (8.7), it follows that the gradient $Df(x)$ of $f(x)$ is

$$Df(x) = -2D(x^t A^t y) + D(x^t A^t Ax), \tag{8.8}$$

since $||y||^2$ is not a function of x and therefore $D(||y||^2) = 0$.

Recall from (10.44) and (10.47) the following gradient formulas:

$$D(x^t C) = C^t \tag{8.9}$$

$$D\left(x^t Mx\right) = 2\,(Mx)^t\,, \tag{8.10}$$

for any constant column vector C and for any symmetric matrix M. Since $A^t A$ is a symmetric matrix, see Lemma 5.2, we obtain from (8.9) and (8.10) that

$$D(x^t A^t y) = (A^t y)^t; \tag{8.11}$$

$$D(x^t A^t Ax) = 2(A^t Ax)^t. \tag{8.12}$$

From (8.8), (8.11), and (8.12), it follows that

$$\begin{aligned} Df(x) &= -2(A^t y)^t + 2(A^t Ax)^t \\ &= 2(A^t Ax - A^t y)^t. \end{aligned} \tag{8.13}$$

From (8.13), we find that $Df(x_0) = 0$ if and only if

$$A^t Ax_0 = A^t y. \tag{8.14}$$

Since the columns of A are linearly independent, it follows from Lemma 5.2 that the matrix $A^t A$ is symmetric positive definite. Then, $A^t A$ is a nonsingular matrix, see Lemma 5.3, and therefore the unique solution of (8.14) is

$$x_0 = (A^t A)^{-1} A^t y. \tag{8.15}$$

To classify the critical point x_0 given by (8.15), we compute the Hessian of $f(x)$. From (10.46) and (10.48), we obtain that $D^2(x^t A^t y) = 0$ and $D^2\left(x^t A^t Ax\right) = 2(A^t A)$. Then, from (8.7), it follows that $D^2 f(x) = 2(A^t A)$, and therefore

$$D^2 f(x_0) = 2(A^t A).$$

Since the matrix A^tA is symmetric positive definite, we obtain that the Hessian $D^2f(x_0)$ is symmetric positive definite, and therefore x_0 is a minimum point for the function $f(x)$. We conclude that the point x_0 is a global minimum point for $f(x)$, since x_0 is the only critical point of $f(x)$.

Thus, the solution to the least squares problem (8.1) is given by (8.15), i.e.,

$$x \ = \ (A^tA)^{-1}A^ty. \tag{8.16}$$

Note that the numerical value of x from (8.16) is computed by solving the linear system $(A^tA)x = A^ty$ using the Cholesky solver from Table 6.2, since A^tA is a symmetric positive definite matrix; see the pseudocode from Table 8.1 for details.

Table 8.1: Least squares implementation

```
Function Call:
x = least_squares(A,y)

Input:
A = m × n matrix; m > n
y = column vector of size m

Output:
x = solution to min ||y − Ax||

x = linear_solve_cholesky(A^t A, A^t y);
```

8.1.1 Least squares for implied volatility computation

Consider a European call or put option[1] on an underlying asset whose price is assumed to follow a lognormal model. The implied volatility of the option is the unique value of the volatility parameter σ from the lognormal model that makes the Black–Scholes value of the option equal to the market price of the option.

More precisely, if C_m and P_m are the market prices of a European call option and of a European put option, respectively, with strike K and maturity T on an underlying asset with spot price S paying dividends continuously at the rate q, and assuming that interest rates are constant and equal to r, the implied volatility σ_{imp} corresponding to the price C_m is, by definition, the solution $\sigma = \sigma_{imp}$ to

$$C_{BS}(S, K, T, \sigma, r, q) \ = \ C_m; \tag{8.17}$$

the implied volatility σ_{imp} corresponding to price P_m is the solution $\sigma = \sigma_{imp}$ to

$$P_{BS}(S, K, T, \sigma, r, q) \ = \ P_m. \tag{8.18}$$

Here, $C_{BS}(S, K, T, \sigma, r, q)$ and $P_{BS}(S, K, T, \sigma, r, q)$ are the Black–Scholes values of a call option and of a put option given by (10.77–10.80), i.e.,

$$C_{BS}(S, K, T, \sigma, r, q) \ = \ Se^{-qT}N(d_1) - Ke^{-rT}N(d_2); \tag{8.19}$$

$$P_{BS}(S, K, T, \sigma, r, q) \ = \ Ke^{-rT}N(-d_2) - Se^{-qT}N(-d_1), \tag{8.20}$$

[1]See Section 10.3 for a brief overview of European options.