Derivation of Least Squares Estimates in Matrix Form - A refresher

1 Linear Regression Model in Matrix Form

In a linear regression model, we express the system as:

$$Y = X\theta + \varepsilon$$

where:

- Y is the $n \times 1$ vector of observations,
- X is the $n \times p$ matrix of regression variables,
- θ is the $p \times 1$ vector of unknown parameters,
- ε is the $n \times 1$ vector of error/noise terms.

1.1 Step 1: Define the Least Squares Objective Function

The least squares approach aims to find θ that minimizes the following cost function:

$$J(\theta) = \|Y - X\theta\|_2^2$$

Expanding the Euclidean norm:

$$J(\theta) = (Y - X\theta)^T (Y - X\theta)$$

Expanding the quadratic expression:

$$J(\theta) = Y^T Y - 2\theta^T X^T Y + \theta^T X^T X\theta$$

1.2 Step 2: Compute the Derivative and Set to Zero

To minimize $J(\theta)$, we take its derivative with respect to θ and set it to zero.

Matrix Differentiation Properties

We use the following differentiation rules:

1. **Derivative of a constant:** The derivative of a term that does not involve θ is zero:
$$\frac{d}{d\theta} Y^T Y = 0$$
2. **Derivative of a linear term:** For a constant matrix B,
$$\frac{d}{d\theta} \theta^T B = B$$
3. **Derivative of a quadratic form:** If θ is a vector, and M is a symmetric matrix,
$$\frac{d}{d\theta} \theta^T M \theta = 2M\theta$$

Applying These Rules

Differentiating $J(\theta)$:

$$\frac{d}{d\theta} J(\theta) = \frac{d}{d\theta} \left(Y^T Y - 2\theta^T X^T Y + \theta^T X^T X \theta \right)$$

$$-2X^T Y + 2X^T X\theta = 0$$

1.3 Step 3: Solve for θ

Rearrange the equation:

$$X^T X\theta = X^T Y$$

If $X^T X$ is invertible, we solve for θ:

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

2 Linear Regression Least Squares Fit - An example

We are given the following dataset:

x	1	2	3	4	5	6	7	8	9	10
y	2.1	2.9	3.8	5.0	5.9	7.1	7.8	9.1	9.9	11.2

We assume a linear polynomial model of order 1 that the form:

$$y = \theta_0 + \theta_1 x + \varepsilon$$

where:

- y is the output variable,
- x is the input variable,
- θ_0 (intercept) and θ_1 (slope) are parameters to be estimated,
- ε represents error/noise.

Least Squares Solution

The least squares estimate is given by:

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

where:

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \\ 1 & 10 \end{bmatrix}$$

and

$$Y = \begin{bmatrix} 2.1 \\ 2.9 \\ 3.8 \\ 5.0 \\ 5.9 \\ 7.1 \\ 7.8 \\ 9.1 \\ 9.9 \\ 11.2 \end{bmatrix}$$

The normal equations for estimating θ_0 and θ_1 are:

$$X^T X \hat{\theta} = X^T Y$$

Solving for $\hat{\theta}$:

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

MATLAB Implementation

The following MATLAB code computes the least squares estimate and plots the linear trend.

```
% Define the data
x = (1:10)';  % Column vector of x values
y = [2.1; 2.9; 3.8; 5.0; 5.9; 7.1; 7.8; 9.1; 9.9; 11.2]; % Observations

% Construct the design matrix (adding a column of ones for the intercept)
X = [ones(length(x), 1), x];
Y=y;

% Compute the least squares solution
%theta_hat = (X' * X) \ (X' * y);
theta_hat = X \ Y;

% Extract the intercept and slope
theta_0 = theta_hat(1); % Intercept
theta_1 = theta_hat(2); % Slope

% Display the estimated parameters
fprintf('Estimated intercept (theta_0): %.4f\n', theta_0);
fprintf('Estimated slope (theta_1): %.4f\n', theta_1);

% Generate predicted values
y_fit = X * theta_hat;

% Plot the data points and fitted trend line
figure;
scatter(x, y, 'bo', 'filled'); % Scatter plot of actual data
hold on;
plot(x, y_fit, 'r-', 'LineWidth', 2); % Fitted trend line
xlabel('x');
ylabel('y');
title('Linear Regression Fit');
legend('Data points', 'Fitted Line');
grid on;
```

The MATLAB code above computes the estimated parameters and produces a plot as shown in Figure 1.

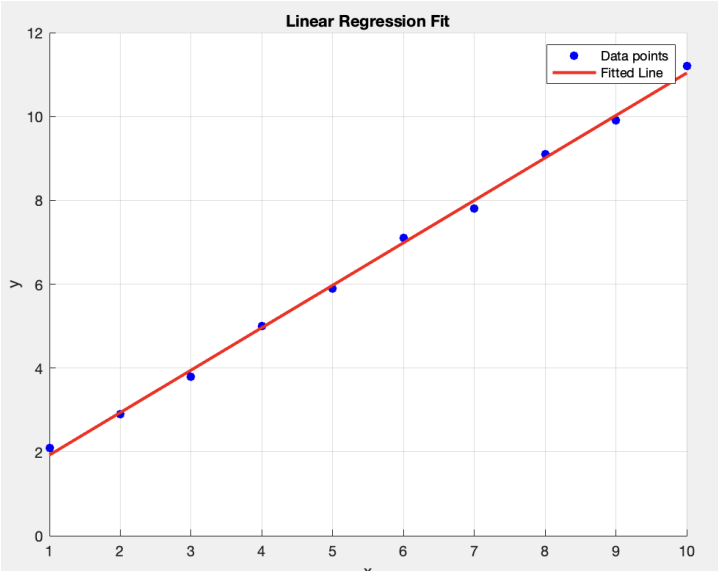

Figure 1: Actual data points and 1st order polynomial model fit estimated by the least squares method

3 Conclusion

The least squares method provides an optimal linear fit for the given data. The estimated parameters θ_0 and θ_1 describe the best-fitting line, minimizing the sum of squared errors.