1. The log-likelihood is then used to make optimization easier:

$$\log L(\beta) = \sum_{i=1}^{n} [y_i \log(P(X_i)) + (1 - y_i) \log(1 - P(X_i))]$$

2. **Gradient Descent or Newton-Raphson Method**: Since the log-likelihood function is non-algebraic (due to the log and exponential terms), it cannot be solved directly. Thus, numerical optimization methods like Newton-Raphson are used to find the β coefficients. These methods iteratively adjust β to maximize the log-likelihood.
3. **Gradient and Hessian**:
   - **Gradient (First Derivative)**: Tells us the direction of steepest ascent in the likelihood function.
   - **Hessian (Second Derivative)**: Tells us the curvature of the function to adjust our steps during optimization.