Lecture Notes:

THE NEWTON-RAPHSON METHOD

July 16, 2024

Jabed Tomal & Hao Xu, Program of Data Science, TRU

Quadratic Model

Example 0.1:

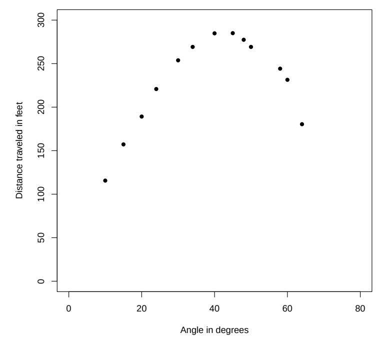
Problem Statement:
We aim to model the distance a baseball travels as a function of the hitting angle using a quadratic equation.
Model Specification:
Let

- Y_i denote the random variable for distance traveled by the baseball (in feet) for which the observed value is y_i.
- X_i denote the random variable of the hitting angle (in degrees) for which the observed value is x_i

We consider that the relationship between the distance travelled and the hitting angle is modeled as the quadratic equation: $Y_i = \alpha + \beta X_i + \gamma X_i^2 + \epsilon_i$,
where

- α, β, γ are the parameters to be estimated, and
- ϵ_i is the error term, assumed to be normally distributed with mean 0 and variance σ^2, i.e., $\epsilon_i \sim N(0, \sigma^2)$.

Objective:
Given a dataset $\{(y_i, x_i)\}_{i=1}^n$, our goal is to estimate the parameters α, β, γ, and σ^2.



Because of $Y_i \sim N(\alpha + \beta X_i + \gamma X_i^2, \sigma^2)$, the probability density function of $Y_i = y_i$ is:

$$P(Y_i = y_i | x_i, \alpha, \beta, \gamma, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2}.$$

Having the observed data, the likelihood function of $(\alpha, \beta, \gamma, \sigma^2)$ can be written as:

$$
\begin{aligned}
L(\alpha,\beta,\gamma,\sigma^2|\{(y_i,x_i)\}_{i=1}^n) &= \prod_{i=1}^n P(Y_i = y_i|x_i,\alpha,\beta,\gamma,\sigma^2) \\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\alpha-\beta x_i-\gamma x_i^2)^2} \\
&= (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\alpha-\beta x_i-\gamma x_i^2)^2}.
\end{aligned}
$$

Taking the log on both sides of the equation yields the following log-likelihood function:

$$
l(\alpha,\beta,\gamma,\sigma^2) = \ln L(\alpha,\beta,\gamma,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \alpha - \beta x_i - \gamma x_i^2). \tag{1}
$$

We want to approximate the value of $(\alpha,\beta,\gamma,\sigma^2)$ by maximizing the likelihood function (Eq. 1).

Newton-Raphson Algorithm:

1. **Gradient Vector**: The vector of first derivatives with respect to the parameters of $l(\alpha,\beta,\gamma,\sigma^2)$ is:

$$
\nabla l(\mathbf{\Theta}) = \begin{pmatrix} \frac{\partial l}{\partial \alpha} \\ \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \gamma} \\ \frac{\partial l}{\partial \sigma^2} \end{pmatrix}, \qquad \text{where } \mathbf{\Theta} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \sigma^2 \end{pmatrix},
$$

and

$$
\begin{aligned}
\frac{\partial l}{\partial \alpha} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i - \gamma x_i^2)(-1) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2); \\
\frac{\partial l}{\partial \beta} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i - \gamma x_i^2)(-x_i) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i)(y_i - \alpha - \beta x_i - \gamma x_i^2) \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i y_i - \alpha\sum_{i=1}^{n} x_i - \beta\sum_{i=1}^{n} x_i^2 - \gamma\sum_{i=1}^{n} x_i^3\right); \\
\frac{\partial l}{\partial \gamma} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(y_i - \alpha - \beta - \gamma x_i^2)(-x_i^2) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i^2)(y_i - \alpha - \beta - \gamma x_i^2) \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i^2 y_i - \alpha\sum_{i=1}^{n} x_i^2 - \beta\sum_{i=1}^{n} x_i^3 - \gamma\sum_{i=1}^{n} x_i^4\right); \\
\frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} - \frac{(-1)(\sigma^2)^{-2}}{2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2 \\
&= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2
\end{aligned}
$$

2. **The Hessian matrix**: The matrix of second derivatives with respect to the parameters of the log-likelihood function is:

$$
\nabla^2 l(\boldsymbol{\Theta}) = \begin{bmatrix}
\frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \beta} & \frac{\partial^2 l}{\partial \alpha \partial \gamma} & \frac{\partial^2 l}{\partial \alpha \partial \sigma^2} \\
\frac{\partial^2 l}{\partial \beta \partial \alpha} & \frac{\partial^2 l}{\partial \beta^2} & \frac{\partial^2 l}{\partial \beta \partial \gamma} & \frac{\partial^2 l}{\partial \beta \partial \sigma^2} \\
\frac{\partial^2 l}{\partial \gamma \partial \alpha} & \frac{\partial^2 l}{\partial \gamma \partial \beta} & \frac{\partial^2 l}{\partial \gamma^2} & \frac{\partial^2 l}{\partial \gamma \partial \sigma^2} \\
\frac{\partial^2 l}{\partial \sigma^2 \partial \alpha} & \frac{\partial^2 l}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 l}{\partial \sigma^2 \partial \gamma} & \frac{\partial^2 l}{(\partial \sigma^2)^2}
\end{bmatrix}
$$

3. Calculation of the second-order partial derivatives yields:

$$
\begin{aligned}
&\nabla^2 l(\boldsymbol{\Theta}) \\
&= \begin{bmatrix}
-\frac{n}{\sigma^2} & -\frac{\sum x_i}{\sigma^2} & -\frac{\sum x_i^2}{\sigma^2} & -\frac{\left(\sum y_i - n\alpha - \beta\sum x_i - \gamma\sum x_i^2\right)}{\sigma^4} \\
-\frac{\sum x_i}{\sigma^2} & -\frac{\sum x_i^2}{\sigma^2} & -\frac{\sum x_i^3}{\sigma^2} & -\frac{\left(\sum x_i y_i - \alpha\sum x_i - \beta\sigma x_i^2 - \gamma\sigma x_i^3\right)}{\sigma^4} \\
-\frac{\sum x_i^2}{\sigma^2} & -\frac{\sum x_i^3}{\sigma^2} & -\frac{\sum x_i^4}{\sigma^2} & -\frac{\left(\sum x_i^2 y_i - \alpha\sum x_i^2 - \beta\sum x_i^3 - \gamma\sum x_i^4\right)}{\sigma^4} \\
-\frac{\left(\sum y_i - n\alpha - \beta\sum x_i - \gamma\sum x_i^2\right)}{\sigma^4} & -\frac{\left(\sum x_i y_i - \alpha\sum x_i - \beta\sigma x_i^2 - \gamma\sigma x_i^3\right)}{\sigma^4} & -\frac{\left(\sum x_i^2 y_i - \alpha\sum x_i^2 - \beta\sum x_i^3 - \gamma\sum x_i^4\right)}{\sigma^4} & \frac{\partial^2 l}{(\partial \sigma^2)^2}
\end{bmatrix}
\end{aligned}
$$

where, the last diagonal element is

$$\begin{aligned}\frac{\partial^2 l}{(\partial\sigma^2)^2} &= -\frac{n}{2}(-1)(\sigma^2)^{-2} + \frac{1}{2}(-2)(\sigma^3)^{-3}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2 \\ &= \frac{n}{2\sigma^4} - \frac{\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2}{\sigma^6}.\end{aligned}$$

4. Finally, the iteration formula for $\mathbf{\Theta}$ is:

$$\mathbf{\Theta}^{\text{new}} = \mathbf{\Theta}^{\text{old}} - \left(\nabla^2 l(\mathbf{\Theta}^{\text{old}})\right)^{-1}\left(\nabla l(\mathbf{\Theta}^{\text{old}})\right)$$

Application:

1. **Drug discovery**: Finding chemical compounds that are active against a disease from a large chemical library.

2. **Protein homology**: Detecting homologous proteins which is useful to find homologous sequence of proteins.

3. **Predicting housing market**: Is government induced policies impact housing prices?

4. **Genetics**: Detecting genes which are responsible for disease/phenotype in human and plant genetics.

5. **Ecology**: Do human induced disturbances have threshold effects on the natural environment for animals, fish, birds and plants?

6. **Public health**: Determining factors that effect public health in countries across the world.