

Linear Models and Least Squares

The Linear model

Given a vector of inputs $\mathbf{x}^T = (x_1, x_2, \dots, x_p)$ we predict the output y with the model

$$\tilde{y} = \beta_0 + \sum_{i=1}^p x_i \beta_i$$

If we include 1 as the first element in \mathbf{x} and collect the parameters β_i in a vector $\boldsymbol{\beta}$ we can write the linear model as an inner product

$$\tilde{y} = \mathbf{x}^T \boldsymbol{\beta}$$

Method of Least Squares

Suppose we have a dataset (\mathbf{x}_i, y_i) for $i = 1, \dots, N$. We can collect the inputs \mathbf{x}_i into an $N \times p$ matrix \mathbf{X} with \mathbf{x}_i^T as the rows

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix}$$

and the outputs y_i into a vector $\mathbf{y}^T = (y_1, y_2, \dots, y_N)$. We pick the coefficients $\boldsymbol{\beta}$ that minimizes the residual sum of squares.

$$\text{RSS}(\boldsymbol{\beta}) = \sum_{i=1}^N \left(y_i - \sum_{j=1}^p X_{ij} \beta_j \right)^2 = \sum_{i=1}^N (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

We do this by taking the gradient of $\text{RSS}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and setting all the components to zero.

$$\begin{aligned}
\frac{\partial}{\partial \beta_k} \text{RSS}(\boldsymbol{\beta}) &= \sum_{i=1}^N \frac{\partial}{\partial \beta_k} \left(y_i - \sum_{j=1}^p X_{ij} \boldsymbol{\beta}_j \right)^2 \\
&= \sum_{i=1}^N 2 \left(y_i - \sum_{j=1}^p X_{ij} \boldsymbol{\beta}_j \right) X_{ik} = 2 \sum_{i=1}^N X_{ki}^T y_i - 2 \sum_{i=1}^N X_{ki}^T \sum_{j=1}^p X_{ij} \boldsymbol{\beta}_j = 0 \\
&\rightarrow \sum_{i=1}^N X_{ki}^T y_i = \sum_{i=1}^N X_{ki}^T \sum_{j=1}^p X_{ij} \boldsymbol{\beta}_j
\end{aligned}$$

In matrix notation this equation reads

$$\mathbf{X}^T \mathbf{y} = \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}$$

which is the condition that $\boldsymbol{\beta}$ must satisfy in order to minimize $\text{RSS}(\boldsymbol{\beta})$. If $\mathbf{X}^T \mathbf{X}$ is not invertible, there are several solutions. If $\mathbf{X}^T \mathbf{X}$ is invertible, the unique solution is given by

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$