Let $\overrightarrow{x_i} = (1, x_{i1}, x_{i2}, \ldots, x_{ip})$ be a datum vector and y_i be its response, $X = [\overrightarrow{x_1}, \overrightarrow{x_2}, \ldots, \overrightarrow{x_n}]^T$ is a matrix containing datum vectors. Let $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ be the coefficients vector to fit and $\vec{\beta}' = (0, \beta_1, \ldots, \beta_p)$ be the coefficients vector without the first coefficient β_0. The ridge regression

$$
\begin{aligned}
R &= \sum_{i=1}^{n} (y_i - (\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}))^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \\
&= \sum_{i=1}^{n} (y_i - \overrightarrow{x_i} \cdot \vec{\beta})^2 + \lambda \vec{\beta}' \cdot \vec{\beta}'
\end{aligned}
$$

To find $\vec{\beta}$ that minimizes $R \rightarrow \nabla_{\vec{\beta}} R = 0$

$$
\begin{aligned}
\nabla_{\vec{\beta}} R &= \sum_{i=1}^{n} \nabla_{\vec{\beta}} (y_i - \overrightarrow{x_i} \cdot \vec{\beta})^2 + \lambda \nabla_{\vec{\beta}} (\vec{\beta}' \cdot \vec{\beta}') \\
&= -2 \sum_{i=1}^{n} \overrightarrow{x_i} (y_i - \overrightarrow{x_i} \cdot \vec{\beta}) + 2\lambda \vec{\beta}' \\
&= -2X \cdot (y_i - X \cdot \vec{\beta}) + 2\lambda \vec{\beta}
\end{aligned}
$$

Using iterations to renew $\vec{\beta}$ so that $\nabla_{\vec{\beta}} R \approx 0$

$$
\vec{\beta}_{new} = \vec{\beta}_{old} - \alpha \nabla_{\vec{\beta}} R = \vec{\beta}_{old} - 2\alpha [\lambda \vec{\beta}' - X \cdot (y_i - X \cdot \vec{\beta})]
$$

where α is the learning rate.