# RSS and the Gradient in Matrix form - with Ridge

Evert Jan Karman

January 1900

Recap: The RSS and the Gradient in matrix form - with Ridge

## 1 The formula for RSS

$$\mathrm{RSS(w)} = (y - Hw)^\top (y - Hw) + \lambda w^\top w$$

## 2 The formula for the Gradient

$$\nabla \mathrm{RSS(w)} = -2H^\top (y - Hw) + 2\lambda w$$

## 3 The formula for RSS with NO penalty for intercept (w0)

$$\mathrm{RSS(w)} = (y - Hw)^\top (y - Hw) + \lambda \sum_{j=1}^{D-1} w_j^2$$

(that is assuming that we have D features and w0 is our intercept)

If we wanted to write this in matrix form, let

$$P = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

and

$$\mathrm{RSS(w)} = (y - Hw)^\top (y - Hw) + \lambda w^\top P w$$

# 4 The formula for the Gradient with NO penalty for intercept (w0)

$$\nabla \mathrm{RSS(w)} = -2H^\top (y - Hw) + 2\lambda \begin{pmatrix} 0 \\ w_1 \\ \vdots \\ w_{D-1} \end{pmatrix}$$

or in matrix form

$$\nabla \mathrm{RSS(w)} = -2H^\top (y - Hw) + 2\lambda P w$$