Sketch of the derivation of the multi-dimensional least squares matrix-vector model:

the observables $\{y_i\}$ now depend "linearly" on more than one i/p features, say two i/p (features) dimensions - namely x_{2i} and x_{3i} (generally we do not use x_{1i} by convention as a symbol in this model)

Data pts (say we have 10 of them)

$i=1$: $y_1 = \beta_1 + \beta_2 x_{21} + \beta_3 x_{31}$

$i=2$: $y_2 = \beta_1 + \beta_2 x_{22} + \beta_3 x_{32}$

⋮

$i=10$: $y_{10} = \beta_1 + \beta_2 x_{210} + \beta_3 x_{310}$

In matrix vector form:

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{10} \end{pmatrix}_{10\times 1} = \begin{pmatrix} 1 & x_{21} & x_{31} \\ 1 & x_{22} & x_{32} \\ \vdots & \vdots & \vdots \\ 1 & x_{210} & x_{310} \end{pmatrix}_{10\times 3} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}_{3\times 1} = A\vec{\beta}$$

But A is not a square matrix;

$\therefore$ A is not invertible

$\therefore$ $\beta = A^{-1}\vec{y}$ is not a feasible calculation.

Essentially, this entails that a straight-line cannot be drawn through each of the ten data points. the line of "best fit" must be captured as $\vec{y} = A\vec{x} + \vec{e}$ where $\vec{e}$ must be as small as possible (minimized) to ensure a "best fit".

$$\vec{e} = (\vec{y} - A\vec{x})$$

$$E = \vec{e}^{\dagger}\vec{e} = (\vec{y} - A\vec{x})^{\dagger}(\vec{y} - A\vec{x}) = \sum_{i=1}^{10} e_i^2$$

Here $\vec{x}$ is the vector $\vec{\beta}$ (model parameters stacked as a vector)

$$\frac{\partial E}{\partial \vec{\beta}} = 0 = \frac{\partial}{\partial \vec{\beta}}\left\{\vec{y}^{\dagger}\vec{y} - \vec{y}^{\dagger}A\vec{\beta} - \vec{\beta}^{\dagger}A^{\dagger}\vec{y} + \vec{\beta}^{\dagger}A^{\dagger}A\vec{\beta}\right\}$$

Note $(AB)^{\dagger} = B^{\dagger}A^{\dagger}$

Do you Know the meaning of this term?

$$\Rightarrow -2A^{\dagger}\vec{y} + 2A^{\dagger}A\vec{\beta} = 0$$

Solving for $\vec{\beta}$,

$$\vec{\beta} = (A^{\dagger}A)^{-1}A^{\dagger}\vec{y}$$

is the least-squares soln. of the regression model.

Question: (i) Can you tell me a condition when $(A^{\dagger}A)$ is not invertible?

(ii) What do we do when $(A^{\dagger}A)^{-1}$ does not exist?