Er zijn twee manieren om regressie te doen.

1. Minimalisatie van de fout.

   Stel we zoeken een lijn door een aantal punten. De lijn is gegeven door

   $$y = ax + b$$

   We zoeken nu a en b zodanig dat de lijn zo goed mogelijk past op de data (meestal zal een rechte lijn niet door alle data-punten heen kunnen).

   Stel we hebben wat punten: $(x_1, y_1), (x_2, y_2)$ etc. We kunnen nu de fout van de lijn definieren als het kwadraat van de afstand tussen de y die de lijn voorspelt op een bepaalde x en de y die daadwerkelijk waargenomen wordt bij die x. Dan krijg je:

   $$E(a, b) = \sum_{i}^{n} (y_i - (ax_i + b))^2$$

   Waar n het aantal gegeven punten is. Een optimum van deze fout is simpel te berekenen. Je neemt de gradient gelijk aan 0 en vindt een a en b. Deze methode werkt natuurlijk ook als we bijvoorbeeld een polominaal zoeken.

   Stel we zoeken tweede orde polominaal:

   $$y = ax^2 + bx + c$$

   Dan wordt de fout:

   $$E(a, b, c) = \sum_{i}^{n} (y_i - (ax_i^2 + bx_i + c))^2$$

2. Pseudo inverse.

   Voor het vinden van een regressielijn of -vlak is nog een andere methode. We weten immers dat als een lijn perfect door de punten zou kunnen je dan hebt:

   $$X\vec{w} = \vec{y}$$

X is hier een matrix met de input-waardes, $\vec{w}$ zijn de gewichten (a en b van hierboven) en $\vec{y}$ is de uitkomst vector. Uitgeschreven voor 2 dimensies en 3 punten wordt dit:

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

De kolom met 1-en is om te zorgen dat de lijn niet perse door de oorsprong gaat (anders zouden we een lijn $y = ax$ ipv $y = ax + b$ aan het zoeken zijn). Nu lijkt dit makkelijk op te lossen, maar als de punten niet op een lijn liggen geeft bovenstaand stelsel geen oplossingen voor a en b. Probeer bijvoorbeeld eens met de punten $(0, 0), (1, 1), (2, 0)$.

Nu blijkt hiervoor een oplossing te bestaan. Onthoud: We weten X en $\vec{y}$, we willen weten $\vec{w}$. Als het mogelijk zou zijn zouden we eigenlijk willen dat je gewoon de inverse van X zou kunnen nemen, want dan had je:

$$\begin{aligned}
&X\vec{w} = \vec{y} \\
&X^{-1}X\vec{w} = X^{-1}\vec{y} \\
&\vec{w} = X^{-1}\vec{y}
\end{aligned}$$

Als je dan de inverse van X hebt, en je hebt $\vec{y}$, dan kun je $\vec{w}$ gewoon uitrekenen. Nu kan je over het algemeen geen inverse nemen, omdat deze niet bestaat[1]. Dan is er wel de mogelijkheid tot het nemen van de pseudo-inverse, als volgt:

$$\begin{aligned}
&X\vec{w} = \vec{y} \\
&(X^TX)\vec{w} = X^T\vec{y} \\
&(X^TX)^{-1}(X^TX)\vec{w} = (X^TX)^{-1}X^T\vec{y} \\
&\vec{w} = (X^TX)^{-1}X^T\vec{y}
\end{aligned}$$

Hier is $(X^TX)^{-1}X^T$ de pseudo inverse van X. Deze wordt ook wel weergegeven als X^+. Het bewijs dat de oplossing die hieruit komt inderdaad de best passende lijn zal ik hier niet geven[2].

Als het goed is komt uit beide methodes dezelfde uitkomst.

[1]Zou hij wel bestaan dan had je sowieso een vierkante matrix, en dat betekent dat je evenveel punten als dimensies had. Door zoveel punten is altijd een lijn, vlak of hyperplane te trekken. Regressie is dan niet meer nodig.

[2]Maar bestaat natuurlijk wel.