The Pseudoinverse and SVD

Solving Least Squares for Any Matrix

Recap: The Least Squares Problem

Given a linear system $Xw = y$, where $X \in \mathbb{R}^{n \times p}$, $w \in \mathbb{R}^p$, and $y \in \mathbb{R}^n$.

The Challenge

If $n > p$ (more equations than unknowns), there is generally no exact solution. The vector y does not lie in the column space of X.

The goal of Least Squares is to find the "best" approximate solution $\hat{w}$ that minimizes the squared norm of the residual vector $r = y - Xw$.

$$\hat{w} = \arg\min_{w \in \mathbb{R}^p} \|Xw - y\|_2^2$$

The Standard Solution: Normal Equations

If X has full column rank ($\text{rank}(X) = p$), the unique solution is given by:

$$\hat{w} = (X^T X)^{-1} X^T y$$

When Do the Normal Equations Fail?

The standard solution relies on the matrix $(X^T X)$ being invertible.

Problem Scenarios

This fails if X is **not** full column rank. This happens when:

- The columns of X are linearly dependent (collinearity in statistics).
- There are more features than data points ($p > n$).

In these cases, $X^T X$ is singular and cannot be inverted.

The Consequence

The least squares problem still has a solution (in fact, infinitely many solutions), but we cannot find it using the standard formula. We need a more general tool.

The Moore-Penrose Pseudoinverse

The pseudoinverse, denoted X^+, is a generalization of the matrix inverse to any matrix, including non-square or singular matrices.

Definition

For any matrix $X \in \mathbb{R}^{n \times p}$, its pseudoinverse $X^+ \in \mathbb{R}^{p \times n}$ is the unique matrix that satisfies the four Penrose conditions:

1. $XX^+X = X$
2. $X^+XX^+ = X^+$
3. $(XX^+)^T = XX^+$ (The projection onto the column space of X is symmetric)
4. $(X^+X)^T = X^+X$ (The projection onto the row space of X is symmetric)

If X is invertible, then $X^+ = X^{-1}$.

Constructing the Pseudoinverse from SVD

The most stable and general way to compute the pseudoinverse is via the Singular Value Decomposition (SVD).

1. **Decompose** X **using SVD:**

$$X = U\Sigma V^T$$

   where U and V are orthogonal, and Σ is a rectangular diagonal matrix of singular values.

2. **Compute the pseudoinverse of** Σ**:** Σ^+ is obtained by transposing Σ and taking the reciprocal of the **non-zero** singular values.
   If $\Sigma_{ii} = \sigma_i \neq 0$, then $(\Sigma^+)_{ii} = 1/\sigma_i$. All other entries are zero.

3. **Reconstruct** X^+**:**

$$X^+ = V\Sigma^+U^T$$

The Least Squares Solution via Pseudoinverse

Using the pseudoinverse, we can state a completely general solution for the least squares problem.

The General Solution

The solution to $\arg\min_w \|Xw - y\|_2^2$ is given by:

$$\hat{w} = X^+ y$$

Why is this solution special?

- **It always exists:** The pseudoinverse is defined for any matrix X.
- **It is the minimum norm solution:** If the least squares problem has infinitely many solutions (i.e., when X is rank-deficient), the vector $\hat{w} = X^+ y$ is the one with the smallest Euclidean norm $\|\hat{w}\|_2$.

Connecting to the Normal Equations

What is the relationship between X^+ and the solution from the normal equations?

The Full Rank Case

If X has full column rank, then it can be shown that:

$$X^+ = (X^T X)^{-1} X^T$$

In this case, the pseudoinverse solution is identical to the normal equation solution:

$$\hat{w} = X^+ y = (X^T X)^{-1} X^T y$$

The pseudoinverse therefore correctly generalizes the familiar formula to all possible cases.

Conclusion: Why SVD is Key

- The SVD provides a powerful theoretical tool to understand and define the pseudoinverse for any matrix.
- It gives us a practical, numerically stable algorithm for computing the least squares solution, especially in ill-conditioned or rank-deficient cases where the normal equations would fail.
- By handling zero or very small singular values correctly, the SVD-based approach avoids the problems of matrix inversion and provides the unique minimum-norm solution.

Summary

The SVD is the foundation upon which the modern, robust solution to the linear least squares problem is built, providing both theoretical insight and a reliable computational method via the pseudoinverse.