University of Tehran
Faculty of New Sciences and Technologies Department of Mechatronics

Practice of Artificial Neural Networks

By:
Omid Moradi

Lecturer:
Dr. Masoudnia

Fall 2024

1) To find the closed-form solution for the MSE loss, we set $X^TX\boldsymbol{\beta} = X^TY$ (slide 14). For a unique solution, X^TX must be invertible. However, in practice, this matrix is often not invertible. Explain why?

In linear regression, we often use the normal equation $X^TX\beta = X^TY$ to solve for β, where X represents the matrix of input features and Y represents the vector of target outputs. For this equation to have a unique solution, X^TX must be invertible. However, in practice, this is not always the case, and there are several reasons why X^TX may not be invertible:

1. **Multicollinearity**: One of the most common reasons is multicollinearity, which occurs when one or more columns of X are linearly dependent. When this happens, X^TX will not have full rank, meaning it will be singular and, therefore, non-invertible. This is because linearly dependent columns do not add new information to the matrix, making it impossible to find a unique solution for β.
2. **Insufficient Data**: If the number of features (columns in X) exceeds the number of observations (rows in X), X^TX will also be singular. This happens because the system is underdetermined, and there isn't enough data to uniquely determine the model parameters. In this situation, there are infinitely many possible solutions for β, which prevents X^TX from being invertible.
3. **Perfectly Correlated Features**: Even if there are enough data points, if two or more features are perfectly correlated, meaning one feature is a linear combination of another, X^TX will not be invertible. Perfect correlation also causes the matrix to lose rank and become singular.

One common way to address this issue is to use regularization techniques, such as Ridge Regression, which adds a small penalty term λI to the matrix X^TX. This ensures that the matrix is invertible by increasing its diagonal elements slightly, making it possible to compute the solution.

Effect of Image Data on Matrix Invertibility

In the context of solving linear regression problems, particularly with image data, the matrix X^TX can often become non-invertible. This happens for a few key reasons:

1. **High Dimensionality in Image Data:** Image data typically involves very highdimensional feature spaces. For instance, a grayscale image of 100×100 pixels contains 10,000 features (one for each pixel), which can lead to a scenario where the number of features (columns) far exceeds the number of observations (rows). When this occurs, the matrix X^TX becomes singular, meaning it is not of full rank and, thus, non-invertible. This happens because there are more features than independent data points, creating dependencies among the columns.

2. **Redundant or Highly Correlated Features in Images:** Image data often contains redundant information, such as neighboring pixels that are highly correlated (e.g., similar intensities in smooth regions of the image). These correlations introduce linear dependencies among the columns of X, making $X^T X$ singular. In such cases, some columns can be expressed as linear combinations of others, which reduces the rank of the matrix and makes it impossible to compute its inverse.

2) Explain the numerical methods for the approximation of the inverse of XTX when it is not invertible. Hint: use ridge regression and the pseudo-inverse in your search

Numerical Methods for Approximating the Inverse of $X^T X$ When It Is Not Invertible

When the matrix $X^T X$ is non-invertible or ill-conditioned, direct computation of its inverse is not possible. However, there are several numerical methods that can be used to approximate the inverse or solve the system without explicitly computing it. Three of the most important methods are:

1. **Moore-Penrose Pseudoinverse**: The Moore-Penrose pseudoinverse is a generalization of the matrix inverse, particularly useful when the matrix $X^T X$ is singular or not of full rank. It allows for solving underdetermined or overdetermined systems. The pseudoinverse is typically computed using Singular Value Decomposition (SVD), which decomposes X into the product of three matrices $X = U\Sigma(V^T)$ The pseudoinverse is obtained by inverting the non-zero singular values in Σ, and setting the rest to zero. This method is useful when dealing with rank-deficient matrices, providing a stable solution in the least squares sense.

2. **Ridge Regression**: Tikhonov regularization, also known as Ridge Regression in statistics, addresses the invertibility issue by adding a regularization term to the normal equation. Specifically, it modifies the normal equation to:

$$\beta = (X^T X + \lambda I)^{-1} X^T Y$$

   where λ is a positive scalar and I is the identity matrix. This regularization term ensures that the matrix $X^T X + \lambda I$ is always invertible, even when $X^T X$ is singular or nearly singular. The regularization also helps reduce the effect of multicollinearity and stabilizes the solution by penalizing large coefficients.

3. **Conjugate Gradient Method**: The Conjugate Gradient (CG) method is an iterative algorithm used to solve systems of the form $A\beta = b$, where $A = X^\wedge T\ X$. It is particularly effective for large, sparse, symmetric, and positive semi-definite matrices. Rather than explicitly computing the inverse, the CG method iteratively refines an initial guess for β, converging to the solution over several iterations. This method is efficient for solving

large systems where direct inversion would be computationally expensive or numerically unstable.