CSCE 689: Special Topics in Modern Algorithms for Data Science Fall 2023

Lecture 30 – November 10, 2023

Prof. Samson Zhou *Scribe: Shima Salehi*

0.1 Overview of Linear Regression

This class extensively covers basic linear algebra concepts and their applications in regression. Regression, a crucial component of supervised learning, involves matrix A, representing n observations with d features, and matrix b, denoting outcome labels for each observation. The primary goal is to find the vector x such that $Ax = b$.

0.2 Motivation for Linear Regression

The discussion explores motivations for using linear regression models, emphasizing their significance in prediction and causal inference. For instance, in healthcare, regression aids in predicting the length of patient hospital stays (prediction) and understanding why patients require prolonged hospitalization (causal inference). Similarly, in economics and finance, linear models are employed to forecast stock prices based on historical data and relevant factors.

0.3 Regression Applications in Various Fields

Regression techniques find extensive applications in diverse fields. They aid in analyzing relationships between factors in medical research, predicting diseases based on risk factors like smoking or diet, and supporting insurance decisions for better patient care. Additionally, in sports analytics, regression models enable the forecasting of player performance or game outcomes based on historical data. Moreover, regression is instrumental in natural sciences for climate modeling to predict changes in temperature or sea level and analyzing relationships between pollution levels and industrial activity.

0.4 Conditions for Solving $Ax = b$

In the class discussion, we dive into understanding the prerequisites for solving equations in the form of $Ax = b$. This equation represents a system where matrix multiplication (A) applied to an unknown vector (x) results in a known outcome vector (b). However, finding x isn't always straightforward.

Consider a scenario where the system is inconsistent, meaning there's no valid solution. Take the example:

$$X_1 + 2X_2 = 1$$
$$X_1 + 3X_2 = 1$$

In this case, the equations lead to contradictory results, showing that there's no solution for X_1 and X_2 that satisfies both equations simultaneously.

Determining the system's consistency involves examining the rank of two critical matrices: the coefficient matrix A and the augmented matrix formed by combining A with vector b. If matrix A is square ($n \times n$) and is full rank (its rank equals its dimension, i.e., rank $A = n$), then A possesses an inverse A^{-1} such that $AA^{-1} = A^{-1}A = I_n$, where I_n is the identity matrix of order n.

For a general matrix A with linearly independent columns ($n \times d$), the Moore-Penrose inverse or pseudoinverse ($A^\dagger$) of A is defined such that $A^\dagger = A^T(AA^T)^{-1}A$ and satisfies $A^\dagger A = I_d$, where I_d is the identity matrix of order d.

When $n = d$, indicating a square matrix, there exists a unique solution, expressed as $x = A^{-1}b$. However, if $n < d$, the system might have infinite solutions. For instance, in the equation $X_1 + X_2 = 1$, there are infinite possible combinations of X_1 and X_2 that satisfy the equation.

In cases where no solution exists for finding x to satisfy $Ax = b$, linear regression comes into play. This regression aims to minimize the error between Ax and b by choosing x that minimizes this discrepancy, typically through least squares optimization using norms like the l_2 norm.

For example, let's consider the matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. In this scenario, A doesn't possess an inverse (A^{-1}), indicating an inconsistent system and thus no valid solution for $Ax = b$.

This understanding underscores the importance of examining the system's properties and structure to determine the feasibility and existence of a solution in solving equations of the form $Ax = b$.

0.5 Linear Regression and Optimization

In the context of $Ax = b$, the objective is to minimize the problem $L(Ax - b)$ for some selected loss function L, where $Ax = b$. Typically, the two-norm $\|Ax - b\|_2$ is employed in least squares optimization to fit a line through a set of points as closely as possible. This method involves minimizing the error, which is highly sensitive to outliers when using the two-norm. The discussion further emphasizes the statistical and mathematical trade-offs between the two-norm and the one-norm.

0.6 Projection and Solution

To address $Ax = b$ where x is unknown, the class delves into projecting b onto the column space of A. It decomposes b into its components $\mathbf{b}_\perp$ and $\mathbf{b}_\|$. Minimizing $||Ax - b||_2^2$ yields unique solutions. The discussion encompasses solutions for both L_2 and L_1 norms, illustrating that a closed-form solution for L_1 is not available.

0.7 Conclusion

While the L_2 norm provides a solution in the form $x = A^\dagger b_\|$, the L_1 norm does not have a closed-form solution.