

Integrated Design of Artificial Intelligence Application Cases for Basic Mathematics Courses in Universities

Sun Yan

School of Applied Science, Beijing Information Science and Technology University, Beijing, China
Email: sunyan@bistu.edu.cn

How to cite this paper: Yan, S. (2026) Integrated Design of Artificial Intelligence Application Cases for Basic Mathematics Courses in Universities. *Journal of Computer and Communications*, 14, 69-78. <https://doi.org/10.4236/jcc.2026.148004>

Received: July 7, 2026

Accepted: August 25, 2026

Published: August 28, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The rapid development of artificial intelligence technology has placed new demands on university-level public foundational mathematics courses. How to organically integrate the core algorithmic concepts of artificial intelligence into the teaching of the three courses—Advanced Mathematics, Linear Algebra, and Statistics—and develop a systematic, integrated design has become an important issue in current mathematics education. Using “machine learning model training” as an example, this paper proposes an integrated design of AI application cases across the three foundational mathematics courses. This approach aims to break down barriers between courses, help students build a cross-course mathematical knowledge network, and enhance their comprehensive ability to apply mathematical tools to solve practical problems in artificial intelligence.

Keywords

Artificial Intelligence, Foundational Mathematics Courses, Integrated Design

1. Introduction

As a strategic technology leading a new round of scientific and technological revolution and industrial transformation, artificial intelligence is profoundly reshaping the content and mode of higher education teaching. In 2018, the Ministry of Education issued the “Action Plan for AI Innovation in Higher Education Institutions”, explicitly calling for “emphasis on the cross-disciplinary integration of AI with mathematics, computer science, physics, biology, psychology, sociology, and other disciplines”. In 2024, the Ministry of Education further launched the “AI Empowering Education Initiative”, using AI technology as an engine to drive

deep transformation in the form of education and teaching [1]. Against this background, how to integrate the core ideas and methods of AI into the teaching of foundational mathematics courses has become an important proposition facing university mathematics educators.

Advanced Mathematics, Linear Algebra, and Statistics are foundational mathematics courses widely offered at universities, particularly at engineering institutions focusing on information-related disciplines. In many universities, these courses are delivered as separate subjects. They are often scheduled in different semesters and taught by different instructors, so students may not always see how ideas from one course relate to those from another [2]. Therefore, we explored whether there are practical cases that can effectively connect the content of these courses. Machine learning provides a useful illustration. It involves the gradient descent algorithm (derivatives and optimization in Advanced Mathematics), matrix computation (matrix multiplication and eigendecomposition in Linear Algebra), and probabilistic modeling of loss functions (maximum likelihood estimation in Probability and Statistics). These three components naturally form a complete knowledge chain.

In recent years, scholars at home and abroad have conducted extensive explorations on the integration of mathematics courses with AI. Li Hang [3] demonstrated the key role of probabilistic and statistical methods in the design of machine learning algorithms in *Statistical Learning Methods*. Goodfellow *et al.* [4] systematically expounded the core roles of linear algebra, probability theory, and numerical computation in deep learning in the book *Deep Learning*, providing rich application materials for mathematics teaching. Domestic scholars Zhao Qiuxia *et al.* [5] explored extended teaching models that integrate AI into Probability and Statistics courses; Zhang Yu *et al.* [6] designed probability and statistics cases based on the concept of curriculum ideology and politics; Yang Wenxia *et al.* [7] empowered Linear Algebra course teaching with knowledge graphs and explored pathways for constructing educational knowledge graphs. However, existing research mostly focuses on the level of a single course, lacking a systematic design that treats the three foundational mathematics courses as an integrated whole and threads a single case throughout all of them.

Based on this, the present paper proposes an integrated design scheme for AI application cases centered on “machine learning model training” as the core case, organically linking three major modules—the gradient descent algorithm (Advanced Mathematics), matrix computation (Linear Algebra), and probabilistic statistical modeling (Probability and Statistics)—to achieve collaborative teaching of the three courses within the same case framework. This design aims to break down course barriers, help students understand the mathematical foundations of AI from a holistic perspective, and enhance their cross-course knowledge transfer ability and comprehensive competence in solving real-world problems.

2. Core Connections between Foundational Mathematics Courses and Artificial Intelligence

The core task of artificial intelligence can be summarized as: given data, find a

function or model that optimally fits the data, and use it for prediction and decision-making. This process naturally involves three types of mathematical tools: characterizing optimization processes with calculus, representing and processing data with linear algebra, and modeling uncertainty with probability and statistics. **Table 1** shows the mapping relationships between some of the typical AI tasks we designed and the core knowledge points of the three mathematics courses.

Table 1. Mapping relationships between typical AI tasks and core knowledge points of foundational mathematics courses (Partial).

Typical AI Tasks	Calculus Core Knowledge Points	Linear Algebra Core Knowledge Points	Probability and Statistics Core Knowledge Points
Linear Regression	Least squares method, gradient descent	Matrix inversion, vector projection	Maximum likelihood estimation, normal distribution, central limit theorem
Neural Networks	Backpropagation (chain rule), vanishing/exploding gradients, derivatives of activation functions	Matrix operations (multiplication, transpose), weight matrices	Batch normalization, Bayesian interpretation
Reinforcement Learning	Policy gradient theorem (log-derivative trick), gradient ascent	Value function vectors, policy parameter matrices, matrix operations under linear approximation	Markov decision process, expected return, Monte Carlo sampling, likelihood ratio

As can be seen from **Table 1**, the complete implementation of an AI task almost always requires simultaneously drawing upon knowledge from all three mathematics courses. For example, training a linear regression model to predict housing prices first requires representing housing feature data with matrices (Linear Algebra), then defining the mean squared error loss function (Probability and Statistics), and finally iteratively solving for the optimal parameters via gradient descent (Advanced Mathematics). The knowledge from the three courses in this process is not simply in a parallel relationship but forms an organic whole that is interdependent and synergistic. However, in traditional discipline-segregated teaching, these three links are split across three semesters and three instructors' classrooms, making it difficult for students to establish the logical connections between them, thus leading to the dilemma of "learning without knowing how to apply".

Therefore, it is necessary to proceed from a unified perspective and design a teaching case that can thread through the core knowledge points of the three courses, so that when studying each course, students can understand the application scenarios of the knowledge learned and its connections with knowledge from other courses within the framework of this case, thereby forming a systematic mathematical knowledge network.

3. Integrated Teaching Design Centered on "Linear Regression" as the Core Case

This section takes the most classic problem in machine learning—the training of linear regression models—as the core case threading through the three courses, elaborating in detail on how to conduct teaching around this case in each course

and achieve knowledge continuity between the three courses.

3.1. Advanced Mathematics Perspective: Optimization Principles of Gradient Descent

In the Advanced Mathematics course, when covering chapters on derivatives, partial derivatives, and extrema of multivariate functions, the gradient descent algorithm is introduced as a core case. First, taking one-dimensional linear regression as an example, given a dataset, the mean squared error (MSE) loss function is defined [8]:

$$J(w, b) = \frac{1}{n} \sum_{i=1}^n (w \cdot y_i - b \cdot x_i)^2. \quad (1)$$

At this point, the instructor can naturally introduce the geometric meaning of the derivative—the rate of change of a function at a given point—as well as the concepts of directional derivatives and gradients. The gradient

$\nabla J = (\partial J / \partial w, \partial J / \partial b)$ indicates the direction of steepest ascent of a function; therefore, moving along the negative gradient direction causes the loss function value to decrease. This application can help students understand a practical use of the important concept of using partial derivatives to find extrema of a function.

Furthermore, the instructor should introduce the concept of learning rate α and the parameter update formulas can be given:

$$w := w - \alpha \cdot \partial J / \partial w, \quad b := b - \alpha \cdot \partial J / \partial b. \quad (2)$$

Using Python's Matplotlib library to plot the iterative trajectory of gradient descent will help students intuitively observe the process of parameters gradually converging from initial values to the optimal solution. In this part, the multivariate function case can be derived, and parameters are generalized to vector form, laying the groundwork for matrix form in the subsequent Linear Algebra course. When the limitations of gradient descent are discussed, the relevant optimization context should first be clarified. For linear least-squares regression, the mean squared error loss function is convex and, more specifically, quadratic with respect to the model parameters. Therefore, convergence to the unique global minimum can be guaranteed under an appropriately selected learning rate, and the issue of local minima does not arise. The discussion may subsequently be extended to general non-convex optimization problems, such as those encountered in deep neural networks, in which multiple local minima and saddle points may present substantial challenges. As supplementary content, stochastic gradient descent (SGD) and momentum-based methods may also be introduced to encourage further student exploration.

3.2. Linear Algebra Perspective: Data Matrix Representation and Computation

In the subsequent Linear Algebra course, the case previously introduced in Advanced Mathematics may be extended from the one-dimensional setting to a multivariate setting when matrix multiplication, matrix inverses, and systems of linear

equations are studied. Specifically, a dataset containing m samples and n features is represented as a matrix $X \in R^{m \times n}$, the target variable as a vector $Y \in R^m$, and the parameters as a vector $\beta \in R^n$, then the linear regression model to be concisely written in matrix form:

$$y = X\beta + \varepsilon. \quad (3)$$

The expressive power of linear-algebraic notation is demonstrated by this concise matrix formulation. It also provides a suitable framework through which the practical significance of matrix multiplication may be understood: the product of a matrix and a vector is interpreted as a linear combination of features, while each row of the resulting vector is associated with the predicted value of an individual sample. From this, two paths for solving optimal parameters β can be introduced—the numerical solution via gradient descent and the analytical solution via the normal equation—corresponding respectively to two different thinking paradigms: iterative optimization (Advanced Mathematics) and matrix inversion (Linear Algebra) [9].

Additionally, when covering the chapter on eigenvalues and eigenvectors, the principal component analysis (PCA) dimensionality reduction method can be introduced as an extended case. PCA performs eigendecomposition on the covariance matrix of centered data, selecting the eigenvectors corresponding to the largest eigenvalues to form a projection matrix, thereby achieving a low-dimensional representation of high-dimensional data. This case not only reinforces the geometric meaning of eigenvalues and eigenvectors but also provides the mathematical foundation for students to understand data preprocessing (dimensionality reduction, denoising) in machine learning. Moreover, the study of PCA can be connected with singular value decomposition (SVD): computationally, PCA is often implemented via SVD of the centered data matrix (rather than explicit eigendecomposition of the covariance matrix), which is numerically more stable. However, the two are not equivalent in all respects—their interpretation depends on whether the data have been centered (PCA requires centering; SVD does not) and on which matrix is being decomposed (the data matrix itself versus the covariance matrix) [10].

3.3. Probability and Statistics Perspective: Probabilistic Interpretation of Loss Functions and Model Evaluation

In the Probability and Statistics course, the linear regression case can be interpreted from a brand-new perspective—probabilistic modeling. When covering maximum likelihood estimation (MLE), the instructor can pose a key question: “Why is mean squared error chosen as the loss function in linear regression, rather than absolute error or other forms?” The answer to this question precisely reveals the deep connection between probability and statistics and machine learning.

Assume that the error term ε in the linear regression model follows an independent and identically distributed normal distribution $N(0, \sigma^2)$. Then, given the parameters β , the likelihood function of observing the data y is derived from

[8]:

$$L(\beta) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - x_i^T \beta)^2}{2\sigma^2}}. \quad (4)$$

By taking the negative logarithm of the likelihood function, the maximization of the likelihood is transformed into the minimization of the sum of squared errors, which is equivalent to the minimization of the mean squared error loss function. Through this derivation, it can be shown that least-squares estimation coincides with maximum likelihood estimation under specific assumptions: the error terms are assumed to be independent and identically distributed according to a normal distribution with zero mean and constant variance (*i.e.*, homoscedasticity). It should be emphasized that, when these assumptions are violated—for example, in the presence of heteroscedasticity or correlated errors—the equivalence between least-squares estimation and maximum likelihood estimation may no longer hold. In such cases, alternative estimation procedures, including weighted least squares and generalized least squares, may be more appropriate.

Through this derivation, the optimization methods studied in Advanced Mathematics, the matrix representations introduced in Linear Algebra, and the parameter-estimation methods considered in Probability and Statistics can be integrated within a unified framework. Moreover, the logical coherence among these mathematical tools, as well as the importance of validating model assumptions, can be illustrated. For model evaluation, statistical measures such as the coefficient of determination R^2 , adjusted R^2 , the F -test, and the t -test may be introduced, through which model quality can be assessed from the perspective of statistical significance. In addition, the relationships and distinctions between traditional statistical inference and modern machine-learning methodologies can be demonstrated through the introduction of training-test set partitioning, cross-validation, and other model-selection methods [10].

4. Implementation Pathway for the Integrated Three-Course Linkage Case

4.1. Progressive Design in the Temporal Dimension

Typically, these courses are offered in the sequence of “Advanced Mathematics → Linear Algebra → Probability and Statistics” in the curriculum. The integrated case design should be developed with this chronological feature, through which the complete structure of the case may be progressively presented. At the Advanced Mathematics stage, the one-dimensional case of linear regression and the concept of the loss function are initially introduced with gradient descent as the core. At this stage, students are not required to fully master matrix computation and probabilistic modeling; they only need to establish the basic concepts of optimization. At the end of this stage, the suspenseful question of “how to efficiently represent and compute when the feature dimension increases” is proposed, and the motivation for learning Linear Algebra is planted.

At the Linear Algebra stage, the linear regression problem is reformulated in matrix form, the normal equation and matrix calculus are taught, and PCA dimensionality reduction is introduced as an extension. At this stage, the instructor can review the matrix form of gradient descent from Advanced Mathematics, enabling students to see the correspondence between different mathematical expressions of the same problem. At the end of this stage, the question of “why choose squared loss rather than other forms” is left to provoke students’ thinking about probabilistic models.

As for the Probability and Statistics stage, the probabilistic basis of the least-squares method may be elucidated through maximum likelihood estimation, while the extension to logistic regression may be used to demonstrate the intrinsic correspondence between the selection of a loss function and the assumed probability distribution. At this stage, a complete mathematical framework is expected to have been established, through which the full workflow of a linear regression model—namely, “data $\rightarrow$ model $\rightarrow$ loss function $\rightarrow$ optimization procedure $\rightarrow$ statistical evaluation”—may be systematically understood and organized into a coherent knowledge structure.

4.2. Modular Articulation in the Content Dimension

To ensure consistency and coherence in the transmission of the case across courses, standardized “interface documentation” should be developed, in which a unified system of notation, data formats, and modes of problem formulation for the core case are specified. For example, X may be consistently used to denote the feature matrix, y the target vector, and J the loss function throughout the three courses, thereby reducing potential cognitive confusion caused by inconsistent notation. In addition, a shared case dataset, such as the Boston Housing dataset or the California Housing dataset, should be established. Through the use of the same dataset in different courses, students may be enabled to observe directly the differences and complementarities among mathematical tools when they are applied to the same problem [10].

Furthermore, information-based teaching platforms, such as the Chaoxing Learning Platform and Zhihuishu Platform, may be utilized for the construction of cross-course knowledge graphs, in which the knowledge points involved in the core case are annotated and linked according to course affiliation and logical relationships. When studying a particular knowledge point, students can intuitively see its connections with related knowledge in other courses through the knowledge graph, realizing a self-directed learning style akin to “following a map to find what one seeks”. The introduction of knowledge graphs can also help teaching teams track students’ learning trajectories, providing data support for teaching improvement [7].

4.3. Collaborative Innovation in Teaching Methods

The implementation of the integrated case design requires corresponding reforms

in teaching methods. First, a cross-course teaching team for foundational public courses should be established, with instructors of the three courses regularly conducting joint teaching and research activities to jointly refine case details and ensure consistent expression and logical coherence of the case across different courses. Second, the project-based learning (PBL) method [11] [12] should be introduced. After the three courses are completed, a practical training project—such as “implementing a linear regression predictor from scratch”—is set up, requiring students to comprehensively apply the knowledge from the three courses to complete the full workflow of data preprocessing, model design, parameter solving, and result evaluation, and to implement it through programming. Of course, after each course, students may also complete only the project content related to that specific course, and only finish the entire project once they have completed all three courses. Third, the use of computational tools such as Python or MATLAB should be integrated into classroom teaching, making abstract mathematical concepts perceptible and knowable through visualization and numerical experiments. For example, Matplotlib can be used to dynamically demonstrate the iterative process of gradient descent, and NumPy can be used to demonstrate the efficiency of matrix operations. This dual-track teaching approach of “theoretical derivation + programming verification” has been explored in prior studies and is expected to enhance students’ learning interest and practical ability [5] [13]; the present paper provides the conceptual design for such an approach, with empirical validation planned as future work. Moreover, the rapid development of generative AI has made programming more accessible to students. Finally, online assessment platforms can be utilized to assign staged programming assignments, providing students with immediate feedback through automatic grading systems, and analyzing teaching weak points based on student response data to achieve data-driven precision teaching.

5. Conclusions

This paper addresses the problems of distinct course barriers and fragmented knowledge connections in the current teaching of foundational mathematics courses in universities, and proposes a conceptual integrated design scheme for AI application cases centered on “linear regression” as the core case. With the linear regression model adopted as the practical case, three major knowledge modules—namely, gradient-descent optimization in Advanced Mathematics, matrix representation and computation in Linear Algebra, and parameter estimation and model evaluation in Probability and Statistics—are systematically integrated. Consequently, a complete instructional sequence for the mathematical foundations of artificial intelligence is established. In the temporal dimension, a spiral progressive approach is adopted to gradually reveal the complete picture of the case; in the content dimension, cross-course coherence is ensured through a unified symbol system and shared datasets; in the teaching method dimension, programming practice and knowledge graph technology are employed to support the

teaching design. It should be noted that this paper presents a conceptual teaching design rather than empirical evidence of teaching effectiveness. Planned evaluation indicators for future classroom implementation include: 1) cross-course knowledge transfer measured through pre- and post-intervention assessments; 2) model interpretation ability evaluated through student explanations of model outputs; 3) programming performance assessed via project-based assignments; and 4) student learning interest and self-efficacy surveyed through validated questionnaires. The authors intend to conduct a controlled implementation study across multiple course sections to quantitatively evaluate the proposed design.

The AI era has imposed new demands on mathematics education while also bringing new opportunities. Seamlessly integrating the core ideas and methods of AI into foundational mathematics teaching and cultivating interdisciplinary talents with solid mathematical skills and AI literacy is an important direction that mathematics educators should continuously explore. It is hoped that the integrated design approach presented in this paper can provide useful reference for university faculty and jointly promote the reform and innovation of foundational mathematics course teaching.

Fund

This paper is supported by the Education and Teaching Reform Project of the Mathematics Research Branch of Beijing Higher Education Association and the Beijing Interdisciplinary Science Association (Grant No. SXJC-2024-024).

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Ministry of Education of the People's Republic of China (2018) Action Plan for AI Innovation in Higher Education Institutions.
http://www.moe.gov.cn/srcsite/A16/s7062/201804/t20180410_332722.html
- [2] Hu, Q.T., Zhang, Y. and Wu, W.Y. (2026) Deep Integration of Artificial Intelligence and Education: Logic, Dilemmas and Pathways. *China Educational Technology*, **6**, 1-6+23. (In Chinese)
- [3] Li, H. (2019) Statistical Learning Methods. 2nd Edition, Tsinghua University Press. (In Chinese)
- [4] Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning. MIT Press.
- [5] Zhao, Q.X., Di, G.H., Hou, W.T., *et al.* (2023) Exploration of Extended Teaching Integrating Artificial Intelligence into Probability and Statistics Courses for New Engineering Disciplines. *Studies in College Mathematics*, **26**, 112-115. (In Chinese)
- [6] Zhang, Y., Jiang, X. and Li, F.F. (2024) Case Design for Probability and Statistics Based on the Concept of Curriculum Ideology and Politics. *College Mathematics*, **40**, 114-122. (In Chinese)
- [7] Yang, W.X., Wang, W.H., He, L., *et al.* (2023) Research and Practice of Knowledge Graph Empowering Smart Education: A Case Study of the "Linear Algebra" Course

- at Wuhan University of Technology. *Research in Higher Education of Engineering*, **6**, 111-117. (In Chinese)
- [8] Strang, G. (2019) Linear Algebra and Learning from Data. Wellesley-Cambridge Press.
 - [9] Jolliffe, I.T. and Cadima, J. (2016) Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **374**, Article 20150202.
<https://doi.org/10.1098/rsta.2015.0202>
 - [10] Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition, Springer.
 - [11] Li, Z., Zhou, D.D. and Wang, Y. (2019) Educational Knowledge Graph from the Perspective of “AI+”: Connotation, Technical Framework and Application Research. *Journal of Distance Education*, **37**, 42-53. (In Chinese)
 - [12] Liu, W.R., Li, W., Lu, C.Y., *et al.* (2024) Exploration of the Construction of the “Control Theory and Engineering Practice” Course Cluster Based on Competence Orientation, Integration of Two Modes, and Virtual-Real Collaboration. *Research in Higher Education of Engineering*, **1**, 58-64. (In Chinese)
 - [13] Yin, J.X. and Wang, X.J. (2023) Construction of Knowledge Graph for Statistics and Data Science and Cultivation of Innovative Talents. *Journal of Education, Renmin University of China*, **2**, 69-79. (In Chinese)