Machine Learning: Assignment #1

Professor Shuang Liang

2252709
Xuanhe Yang

April, 2025

Problem 1

Problem 1: Supervised and Unsupervised Learning
Supervised Learning:

- **Definition:** Supervised learning is the task of learning a function that maps an input to an output based on example input-output pairs. The algorithm learns from labeled training data.
- **Working Principle:**
  1. Provide training data with labeled examples (input features X and target outputs Y)
  2. Algorithm learns the mapping function f(X) = Y
  3. Use the learned function to predict outputs for new data
- **Representative Algorithms:**
  - Linear Regression: for predicting continuous values
  - Logistic Regression: for binary classification problems
  - Support Vector Machines (SVM): for classification and regression
  - Decision Trees, Random Forests: for classification and regression
  - Neural Networks: for complex pattern recognition
- **Application Scenarios:**
  - Classification problems: spam detection, image recognition
  - Regression problems: house price prediction, sales forecasting

Unsupervised Learning:

- **Definition:** Unsupervised learning is the task of inferring patterns or structures from data without labeled responses. The algorithm learns from unlabeled data.
- **Working Principle:**
  1. Provide unlabeled data (only input features X)
  2. Algorithm identifies patterns, groupings, or structures in the data
  3. Organize or transform the data based on discovered patterns
- **Representative Algorithms:**
  - K-means Clustering: divides data into K clusters
  - Hierarchical Clustering: creates a tree-like structure of clusters
  - Principal Component Analysis (PCA): dimensionality reduction technique
  - DBSCAN: density-based clustering
  - Autoencoders: for feature learning and data compression
- **Application Scenarios:**
  - Clustering analysis: customer segmentation, document categorization
  - Dimensionality reduction: feature extraction, data visualization
  - Anomaly detection: identifying unusual transactions or behaviors

Problem 2

Problem 2: Mean Squared Error using Leave-One-Out Cross-Validation
Leave-One-Out Cross-Validation Steps:
Round 1: Leave out (0,2) for testing

1. Training set: (2,2), (3,1)
2. Calculate linear regression parameters:
   - Slope w: $w = \frac{y_2 - y_1}{x_2 - x_1} = \frac{1-2}{3-2} = -1$
   - Intercept b: $b = y_1 - w \cdot x_1 = 2 - (-1) \cdot 2 = 2 + 2 = 4$
   - Model: $y = -1x + 4$
3. Predict at test point (0,2): $y_{pred} = -1 \cdot 0 + 4 = 4$
4. Calculate error: $(y_{pred} - y_{true})^2 = (4 - 2)^2 = 4$

Round 2: Leave out (2,2) for testing

1. Training set: (0,2), (3,1)
2. Calculate linear regression parameters:
   - Slope w: $w = \frac{y_2 - y_1}{x_2 - x_1} = \frac{1-2}{3-0} = -\frac{1}{3}$
   - Intercept b: $b = y_1 - w \cdot x_1 = 2 - (-\frac{1}{3}) \cdot 0 = 2$
   - Model: $y = -\frac{1}{3}x + 2$
3. Predict at test point (2,2): $y_{pred} = -\frac{1}{3} \cdot 2 + 2 = 2 - \frac{2}{3} = \frac{4}{3}$
4. Calculate error: $(y_{pred} - y_{true})^2 = (\frac{4}{3} - 2)^2 = (-\frac{2}{3})^2 = \frac{4}{9}$

Round 3: Leave out (3,1) for testing

1. Training set: (0,2), (2,2)
2. Calculate linear regression parameters:
   - Slope w: $w = \frac{y_2 - y_1}{x_2 - x_1} = \frac{2-2}{2-0} = 0$
   - Intercept b: $b = y_1 = 2$
   - Model: $y = 0x + 2$
3. Predict at test point (3,1): $y_{pred} = 0 \cdot 3 + 2 = 2$
4. Calculate error: $(y_{pred} - y_{true})^2 = (2 - 1)^2 = 1$

Final MSE Calculation:

$$MSE = \frac{4 + \frac{4}{9} + 1}{3} \qquad = \frac{\frac{36}{9} + \frac{4}{9} + \frac{9}{9}}{3} = \frac{49}{27} \qquad \approx 1.815$$

Therefore, the MSE obtained using Leave-One-Out cross-validation is $\frac{49}{27} \approx 1.815$.

Problem 3

Problem 3: MAE and MSE Calculation
MAE (Mean Absolute Error) Calculation Steps:
Calculate the absolute error for each month: $|FD - AD|$

Month 1:$|44 - 42| = 2$ Month 2: $|46 - 45| = 1$ Month 3:$|48 - 49| = 1$ Month 4: $|50 - 55| = 5$ Month 5:$|55 - 57| =$

Calculate the average of all absolute errors:

$$MAE = \frac{2+1+1+5+2+0+2+2+1+2+2+2}{12} = \frac{22}{12} = 1.8333$$

MSE (Mean Squared Error) Calculation Steps:
Calculate the squared error for each month: $(FD - AD)^2$

Month 1:$(44 - 42)^2 = 4$ Month 2: $(46 - 45)^2 = 1$ Month 3:$(48 - 49)^2 = 1$ Month 4: $(50 - 55)^2 = 25$ Month 5:$(55$

Calculate the average of all squared errors:

$$MSE = \frac{4+1+1+25+4+0+4+4+1+4+4+4}{12} = \frac{56}{12} = 4.6667$$

Therefore, the MAE is 1.8333 and the MSE is 4.6667.

Problem 4

Problem 4: Confusion Matrix Analysis
1. Calculate the precision and recall for each class:
Class A:

- True Positives (TP_A) = 40 (predicted A and actually A)
- False Positives (FP_A) = 35 + 0 = 35 (predicted A but not actually A)
- False Negatives (FN_A) = 20 + 10 = 30 (actually A but not predicted A)
- Precision = $\frac{TP_A}{TP_A+FP_A} = \frac{40}{40+35} = \frac{40}{75} = \frac{8}{15} \approx 0.5333$
- Recall = $\frac{TP_A}{TP_A+FN_A} = \frac{40}{40+30} = \frac{40}{70} = \frac{4}{7} \approx 0.5714$

Class B:

- True Positives (TP_B) = 85 (predicted B and actually B)
- False Positives (FP_B) = 20 + 10 = 30 (predicted B but not actually B)
- False Negatives (FN_B) = 35 + 40 = 75 (actually B but not predicted B)
- Precision = $\frac{TP_B}{TP_B+FP_B} = \frac{85}{85+30} = \frac{85}{115} = \frac{17}{23} \approx 0.7391$
- Recall = $\frac{TP_B}{TP_B+FN_B} = \frac{85}{85+75} = \frac{85}{160} = \frac{17}{32} \approx 0.5313$

Class C:

- True Positives (TP_C) = 20 (predicted C and actually C)
- False Positives (FP_C) = 10 + 40 = 50 (predicted C but not actually C)
- False Negatives (FN_C) = 0 + 10 = 10 (actually C but not predicted C)
- Precision = $\frac{TP_C}{TP_C+FP_C} = \frac{20}{20+50} = \frac{20}{70} = \frac{2}{7} \approx 0.2857$

Problem 4 continued on next page...

- Recall $= \frac{TP_C}{TP_C+FN_C} = \frac{20}{20+10} = \frac{20}{30} = \frac{2}{3} \approx 0.6667$

2. Calculate macro-average and weighted-average precision and recall:
Macro-average: Macro-average takes the arithmetic mean of each metric across all classes, regardless of class size:

- Macro-average Precision $= \frac{Precision_A+Precision_B+Precision_C}{3} = \frac{\frac{8}{15}+\frac{17}{23}+\frac{2}{7}}{3} \approx 0.5194$
- Macro-average Recall $= \frac{Recall_A+Recall_B+Recall_C}{3} = \frac{\frac{4}{7}+\frac{17}{32}+\frac{2}{3}}{3} \approx 0.5898$

Weighted-average: Weighted-average considers the number of samples in each class:

- Class A samples: 70
- Class B samples: 160
- Class C samples: 30
- Total samples: 260
- Weighted-average Precision $= \frac{Precision_A \times 70+Precision_B \times 160+Precision_C \times 30}{260} = \frac{\frac{8}{15} \times 70+\frac{17}{23} \times 160+\frac{2}{7} \times 30}{260} \approx 0.6315$
- Weighted-average Recall $= \frac{Recall_A \times 70+Recall_B \times 160+Recall_C \times 30}{260} = \frac{\frac{4}{7} \times 70+\frac{17}{32} \times 160+\frac{2}{3} \times 30}{260} \approx 0.5577$

Final Results (to 4 decimal places):

- Macro-average precision: 0.5194
- Macro-average recall: 0.5898
- Weighted-average precision: 0.6315
- Weighted-average recall: 0.5577

Problem 5

Problem 5: Implementation of a Practical Agent on Tongji AI
For this assignment, I have implemented a "Kitchen Assistant" (厨房小助手) agent on the Tongji AI platform that provides practical value to users in their daily lives.

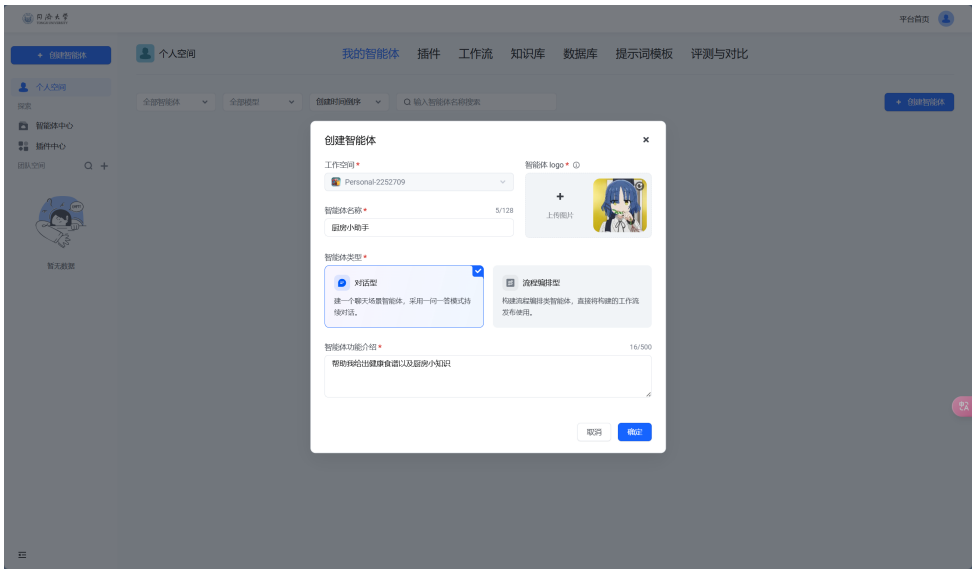

Image 1: Kitchen Assistant agent interface showing recipe instructions

Introduction to the Kitchen Assistant's Functions:
The Kitchen Assistant is designed to offer comprehensive support for cooking and nutrition-related queries. Its primary functions include:

1. **Recipe Recommendations**: The agent can provide step-by-step cooking instructions for various dishes. As shown in Figure 1, it can explain detailed cooking procedures, such as the preparation of tomato and egg stir-fry.

2. **Nutritional Guidance**: The agent offers personalized dietary advice based on users' health goals. It can create balanced meal plans considering nutritional requirements, as demonstrated in its recommendation of a balanced daily meal plan including breakfast, lunch, and dinner options rich in essential nutrients.

3. **Health-Focused Food Knowledge**: The agent possesses knowledge about food properties and their health benefits, helping users make informed dietary choices. For example, it mentions omega-3 fatty acids and various vitamins in its dietary recommendations.

4. **Customized Dietary Suggestions**: The agent can tailor food recommendations to individual needs and preferences, addressing specific health concerns or dietary restrictions.

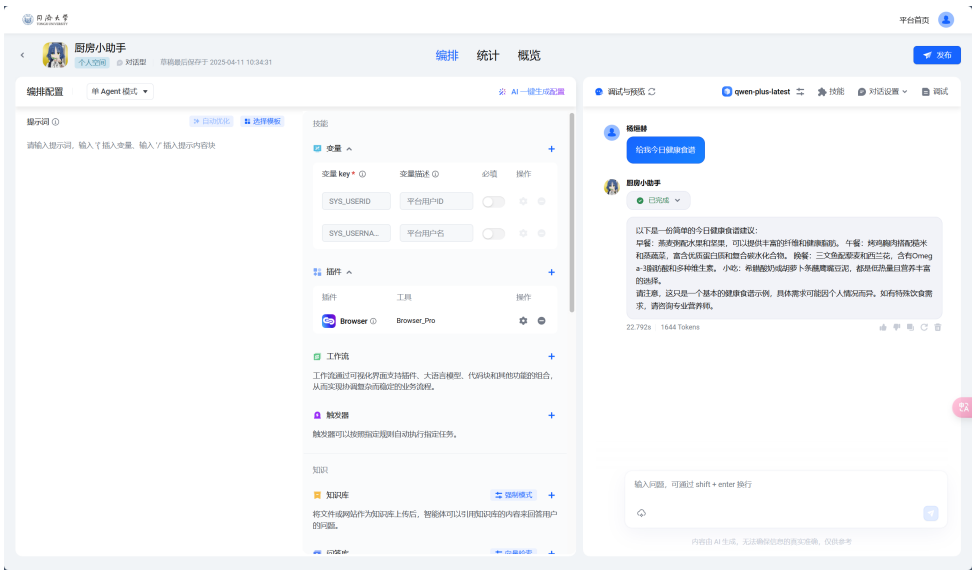

Image 2: Agent creation interface on Tongji AI platform

Implementation Process:
The implementation of this agent involved several key steps:

- **Agent Creation**: As shown in Figure 1, I created the agent in the Tongji AI platform by selecting the "Dialogue Model" (对话型) type, which enables natural conversation flow.

- **Agent Configuration**: I specified the agent's name as "厨房小助手" (Kitchen Assistant) and provided a functional description focused on culinary and nutritional guidance.

- **Knowledge Base Setup**: The agent was configured to access nutrition and cooking information to provide accurate responses to user queries.

- **Conversation Flow Design**: I designed the agent to maintain context in conversations, allowing it to provide coherent responses across multiple interactions, as evident in the conversation screenshots.

Problem 5 continued on next page...

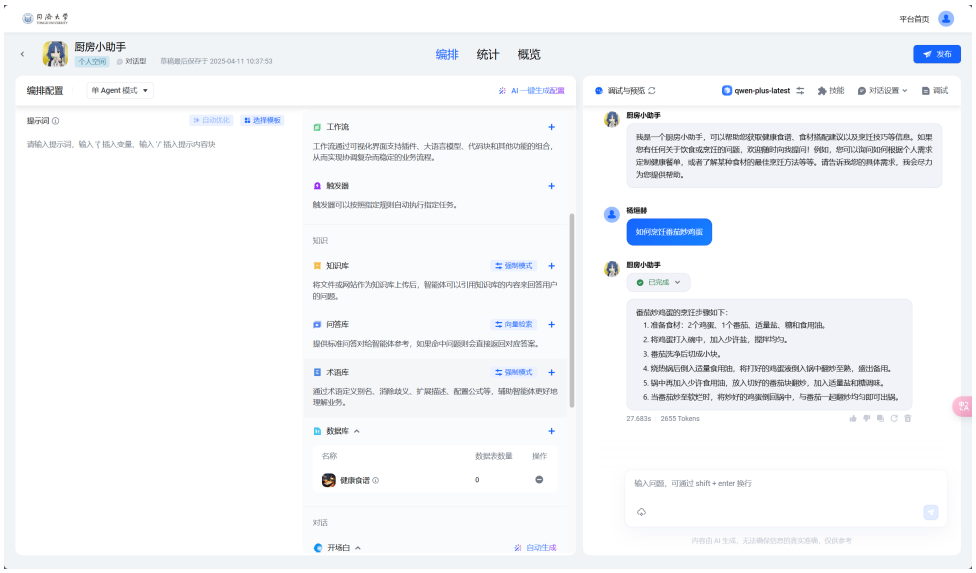

Image 3: Dietary recommendation conversation with the Kitchen Assistant

Application Value:
The Kitchen Assistant offers significant practical value by:

- Helping users prepare healthy meals with detailed instructions
- Providing nutritional education to improve dietary habits
- Offering personalized recommendations based on individual health requirements
- Serving as an accessible culinary resource for students and faculty who may have limited cooking knowledge or time to research recipes

As demonstrated in Figure 1 and Figure 3, the agent can successfully engage in natural-language conversations about food preparation and nutrition, providing specific and detailed information that would be valuable for daily meal planning and preparation.
The implementation demonstrates the practical application of AI conversational agents in supporting everyday activities like cooking and nutrition management, showcasing how these technologies can be integrated into university platforms to enhance student and faculty well-being.