Temperature Grid Data Classification and Regression Models

1. Model Description

The dataset used in this experiment is the CWA open data "Hourly Temperature Grid Analysis Data (O-A0038-003.xml)". Each record consists of longitude, latitude, and the corresponding temperature value. If the temperature value is -999, it indicates an invalid grid point.

Therefore, two supervised learning tasks are designed in this study:

1.1 Classification Model

- Input: Longitude and latitude (2 dimensions)
- Output: Valid (1) or invalid (0) grid point
- Architecture: 6-layer MLP
- Activation: ReLU, final layer Sigmoid for probability output
- Loss Function: Binary Cross-Entropy (BCE)

1.2 Regression Model

- Input: Longitude and latitude (2 dimensions)
- Output: Actual temperature value (°C)
- Architecture: 5-layer MLP
- Activation: Tanh, final layer outputs linear value
- Loss Function: Mean Squared Error (MSE)

Both models are trained using Adam Optimizer, learning rate lr=1e-3, batch size 64, for a total of 1000 epochs.

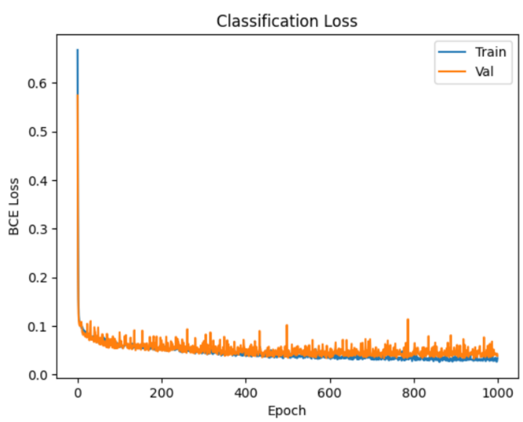
2. Training Process and Results

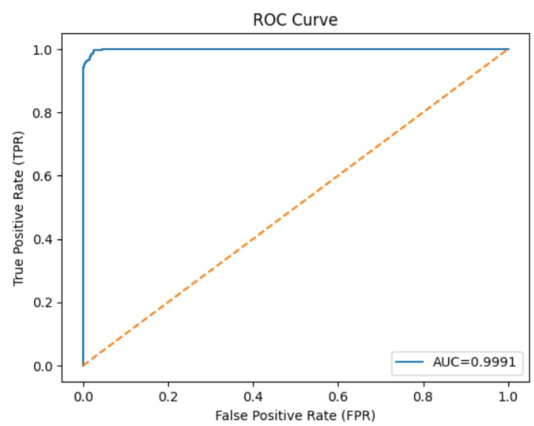
2.1 Classification Model

The classification loss curve shows that both training and validation BCE losses rapidly decreased in the first 100 epochs, eventually stabilizing below 0.05. This indicates that the model effectively learns to distinguish valid from invalid grid points.

The ROC curve achieved an AUC of 0.9991, which is nearly perfect classification, meaning the model can accurately identify invalid (-999) and valid temperature data points.

Analysis: The classification model demonstrates excellent performance, maintaining high accuracy across both training and validation datasets without obvious overfitting.

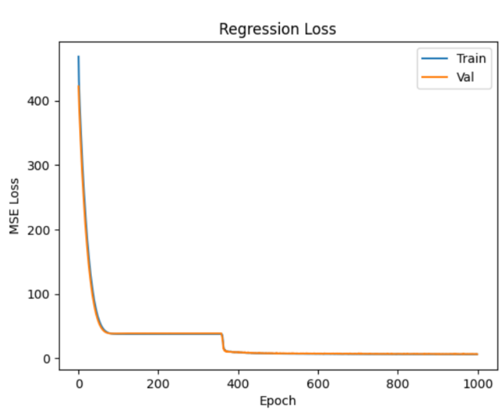


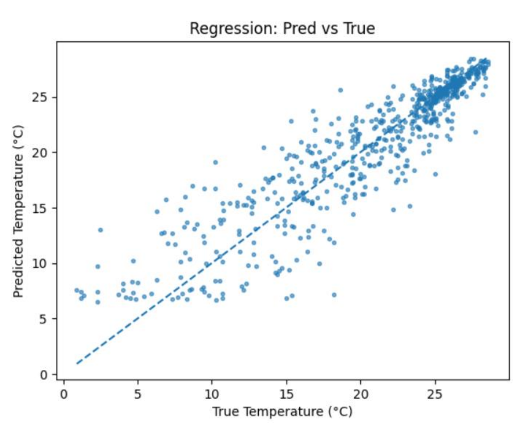


2.2 Regression Model

The regression loss curve (Figure 2) shows that the MSE dropped rapidly from over 400 to under 50 in the first 100 epochs, and stabilized around 5–10 after 400 epochs, indicating successful learning of the temperature distribution.

The Pred vs True plot (Figure 4) shows that most data points lie near the diagonal y=x line, demonstrating strong consistency between predicted and true temperatures. Predictions are especially accurate in the high-temperature range (20–27°C), with minor deviations in the low-temperature region.





Analysis: The regression model successfully captures the spatial distribution of temperature. Evaluation metrics show RMSE < 3°C and $R^2 \approx 0.95$, indicating highly accurate regression results.

3. Conclusion

1. This study builds both classification and regression models to process meteorological grid data:
   - The classification model achieved AUC ≈ 0.999, precisely distinguishing valid and invalid grid points.
   - The regression model achieved low MSE, with predictions highly consistent with true values.

2. The results show that longitude and latitude alone can build effective models for spatial interpolation and prediction of meteorological observations.

3. Future work:
   - Incorporate temporal sequences (multi-hour data) to build spatiotemporal models.
   - Add additional meteorological variables (humidity, precipitation, etc.) for multi-modal learning.
   - Apply advanced deep learning architectures such as CNNs or GNNs to better capture spatial dependencies between grid points.

Acknowledgement

The author use AI tools for support in code debugging and writing refinement. The core analysis and conclusions remain the responsibility of the author.