Estimating Predictive Uncertainty in Neural Networks using Monte Carlo Dropout

Abstract :

Deep Learning models often produce hughly confident predictions even when they are incorrect. In safety-critical domains such as healthcare, understanding predictive uncertainity is essential. This project investigates the use of Monte Carlo Dropout as a simple and effective method to estimate uncertainty in neural network predictions.

1. **Introduction**

   Standard neural networks provide point estimates without indicating confidence. This limitation can lead to overconfident decisions. Monte Carlo Dropout approximates Bayesian inference by enabling dropout during inference, allowing uncertainty estimation through multiple stochastic forward passes.

2. **Methodology**

   A feedforward neural network was trained on a healthcare risk prediction dataset. Dropout layers were included during training and explicitly activated during inference. Multiple forward passes were performed to obtain a distribution of predictions for each sample. Predictive uncertainty was measured as the variance across these predictions.

3. **Experiments**

   The model was evaluated using both standard deterministic predictions and Monte Carlo averaged predictions. Predictive uncertainty distributions were analyzed, and uncertainty levels were compared between correct and incorrect predictions.

4. **Results**

   The results show that incorrect predictions tend to exhibit higher predictive uncertainty than correct predictions. This suggests that Monte Carlo Dropout can provide meaningful uncertainty estimates that correlate with model errors.

5. **Discussion**

   While Monte Carlo Dropout is computationally inexpensive and easy to implement, it provides only an approximation of Bayesian uncertainty. Nevertheless, it offers valuable insights into model reliability and can support safer decision-making.

6. Conclusion

This study demonstrates that uncertainty estimation using Monte Carlo Dropout is a practical approach for assessing prediction reliability in neural networks. Future work may explore more advanced Bayesian methods and calibration techniques.