

Technique for Dimensionality Reduction in Decision Modeling

Marcel Gietzmann-Sanders, Michael Courtney, Andrew
Seitz, Curry Cunningham

University of Alaska Fairbanks

Introduction

Value of Probabilistic Modeling

Theory

Standard probabilistic deep learning networks are typically framed as a classification problem, using categorical cross-entropy as the loss function (Oliver Durr (2020)). Each output neuron represents a potential choice, with the model predicting the probability of each choice being correct based on this loss formulation. For these choices, we provide the network with features encapsulating the relevant information. Training is then comprised of providing a series such decisions.

However, this formulation introduces a critical challenge: if there are N features per choice and M potential choices, the overall dimensionality of the input space becomes $N \cdot M$. Adding even a single feature increases the dimensionality by M not just 1.

This growth poses a significant challenge due to the "curse of dimensionality", where the amount of data required to effectively train models grows exponentially with the dimensionality of the input space (CITE).

Log-Odds Modeling

To address this issue, we propose an alternative framing. Instead of predicting the probabilities directly, we predict the log-odds ϕ_m for each choice and calculate the probability p_m using the softmax function:

$$p_m = \frac{e^{\phi_m}}{\sum_{m=1}^M e^{\phi_m}}$$

This approach reduces the feature space dimensionality to N and effectively increases the number of training examples by a factor of M .

We can implement this log-odds model using standard probabilistic deep learning techniques by replicating the log-odds computation across all M choices. The outputs are fed into a softmax layer with M units, where the layer’s weights are set to the identity matrix and biases are set to zero. Using categorical cross-entropy as the loss function ensures compatibility with standard probabilistic deep learning while enabling us to train the log-odds weights and significantly reduce the problem’s dimensionality.

Contrast Sampling

As M grows large, a practical issue arises: for each training example, most instances of the internal log-odds model would ideally report very low log-odds, resulting in low probabilities. Ideally, only one choice should produce $p_m = 1$. This is analogous to a class imbalance problem, where the model becomes prone to predicting the most common class (CITE).

To address this, we balance the training data. Instead of presenting the model with full decisions containing all M choices, we create training pairs, or contrasts, where each pair consists of one selected choice and one unselected choice. This approach is valid because the log-odds model focuses on the relative likelihood of choices, making the number of choices considered at any one time irrelevant.

The primary risk in using contrasts is introducing bias by disproportionately sampling certain combinations of choices. To mitigate this, we randomly sample pairs from each decision, ensure an equal number of contrasts per decision, and an equal number of decisions per individual. This preserves the balance across the training data and avoids skewing the model’s predictions.

Application

We consider the tracks of 111 Chinook salmon (*Oncorhynchus tshawytscha*)

Data

Model

Training

Results

Discussion

References

Oliver Durr, Beate Sick, E. M. (2020). *Probabilistic Deep Learning*. Manning Publications.