

NEW YORK TAXI FARE CHALLENGE

Till Bungert

ABSTRACT

1 INTRODUCTION

2 RELATED WORK

3 METHODS

3.1 ENTITY EMBEDDINGS

We map categorical variables with C categories represented by indices $c \in [0, C)$ to real-numbered vectors $\mathbf{x}_c \in \mathbb{R}^n$

$$Embedding : [0, C) \rightarrow \mathbb{R}^n, c \mapsto Embedding(c) = \mathbf{x}_c. \quad (1)$$

These embedding layers are implemented as lookup tables. The vector associated with each index is a parameter of the model and is learned jointly with the rest of the model.

If the input to our model is a mixture of continuous and categorical variables as is the case here, we learn one embedding layer for each of the categorical variables and concatenate the vector components of each embedding output together with the continuous variables to one vector. This concatenated vector then serves as the input to the rest of the model.

3.2 DEEP NEURAL NETWORKS

3.3 TREE-BASED METHODS

3.4 BAYESIAN OPTIMIZATION

4 EXPERIMENTS

4.1 DATA SET

5 CONCLUSION