



Input layer consists of all unique course ids. Course ids are assigned embedding ids, and mapped to their unique course id index.

The embedding layer maps individual course ids to a paired context course. The output pairs are assigned a value, and the larger the value, the greater the chances the course is in context. Embedding size = length of embedding ids.

The dense output layer is a softmax which normalizes the outputs with probabilities. Size of dense = length of embedding ids.

$$y_i = \frac{x_i \text{ assigned value from embedding}}{\sum_{j \in \text{Course ids}} y_j}$$

Loss is the sum of the categorical cross entropy for every course in course context.

$$C = - \sum_{j \in \text{Courses in context}} \log y_j$$