

# Learning Label Embeddings for Nearest-Neighbor Multi-class Classification with an Application to Speech Recognition

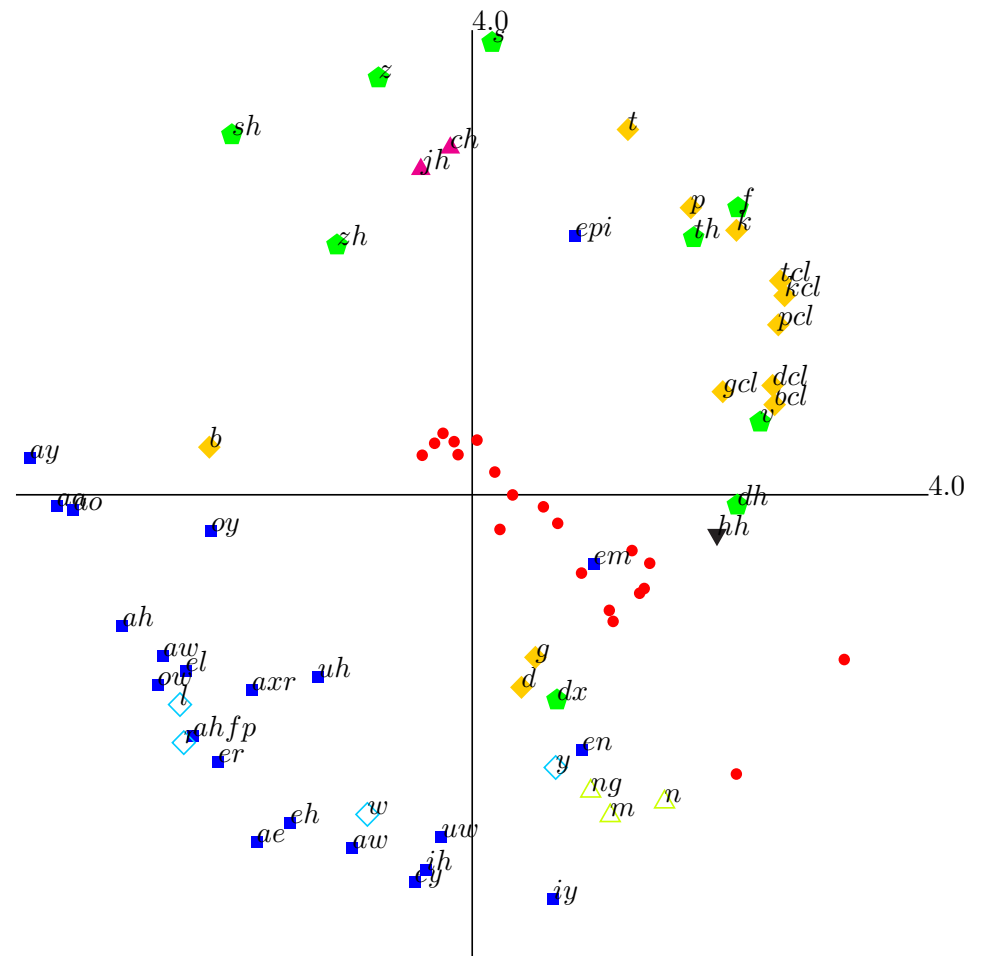
Natasha Singh-Miller and Michael Collins  
Massachusetts Institute of Technology, CSAIL

**Goal:** model  $p(y|a)$  using a nearest neighbor classifier when the number of labels  $y$  is very large

**Example Application:** speech recognition

**Method:** learn an embedding  $y \rightarrow M_y$  where  $M_y \cdot M_z$  represents the similarity of labels  $y$  and  $z$

**Optimization:**  $M_y$  learned to optimize log-likelihood under a leave-one-out criterion



2-D embedding of 73 acoustic-phonetic labels.