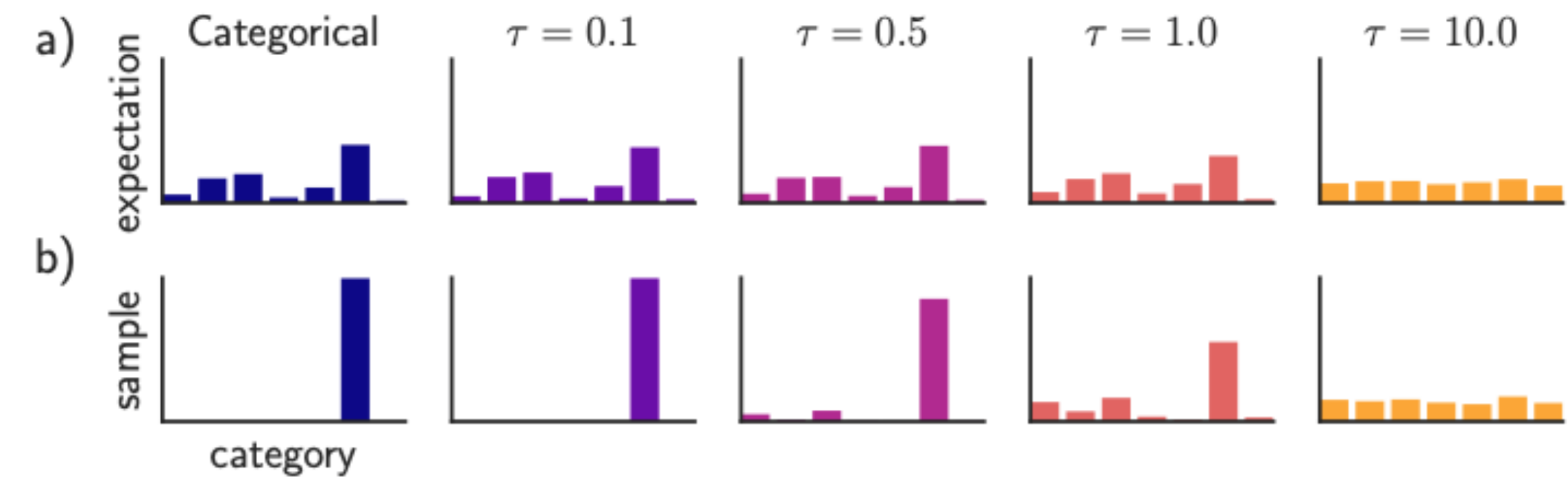


Methodology

Aggregator: Hard-Attention



<https://arxiv.org/abs/1611.01144>

- To overcome this limitation, propose to select the most related context data point instead of weighted average.
- To enable argmax operation to be differentiable, use **Straight-Through (ST) Gumbel SoftMax** (ICLR'17) for **discretely sampling the context information** given target data.
- Through gumbel-softmax, the hard-attention is able to **trim the irrelevant data** and **draw the most informative sample** for given target sample $x_{e,i}$.
- The selected data point $\mathbf{c}_{e,k} \oplus \mathbf{v}_{e,k}$ is fed into fully connected layer that top of the aggregator to adjust dimension and output context embedding $\mathbf{r}_{e,i}$.



Methodology

Detector based on Label Embedding

- Existing works like CNP and ANP usually **simply concat the input feature and numerical label values** together as input.
- These works discard the fact that **label variables are categorical**, and underestimate the importance of labels as dimension of input features is usually significantly larger than single dimensional numerical value.
- Propose to embed labels into fixed dimension vector **inspired by word embedding**.

