Low-Rank Softmax Can Have Unargmaxable Classes in Theory but Rarely in Practice

Andreas Grivas, Nikolay Bogoychev and Adam Lopez

Institute for Language, Cognition, and Computation, University of Edinburgh



A Softmax layer is low-rank when the parametrisation W imposes dimensionality constraints: the number of features d is less that the number of classes n.

Previous work: When d < n - 1

Classifier has reduced expressivity (Softmax Bottleneck*).
Some classes impossible to predict (Stolen Probability[†]).

in Theory

Partition feature space to regions corresponding to rankings.
 Class a is uppergrowable if no region exists for which a is

Feature x_1

Illustration of an *unargmaxable* class. Class c_4 can never be predicted using argmax for this Softmax classifier with n = 4 classes and d = 2 input features. On the left, each feature vector \mathbf{x} is colored according to the class assigned the largest probability; note that while c_1 , c_2 and c_3 surface as regions, c_4 does not. On the right, we show that there is no direction in feature space for which c_4 has the largest probability.

Un·argmax·able Adjective

A class that is impossible to predict irrespective of input.

Rarely in Practice

13/150 models have unargmaxable tokens

FAIR Edinburgh Bergamot Helsinki

Class c_t is unargmaxable if no region exists for which c_t is ranked above all other classes.

Algorithm

Fast ApproximateSlower Exact





	LLMs		Lumburgh	Dergamot	
		MT	MT	MT	MT
unargmaxable	0/7	0/4	0/82	0/25	13 / 32

Number of models found to have unargmaxable tokens. Large Language Models (LLMs) include BERT, (XLM)RoBERT-a and GPT-2. Machine Translation (MT) models checked were obtained from 4 sources.

Unargmaxable tokens are rare and noisy

427 unique unargmaxable BPE tokens.
Rare tokens Предварительны (en-ru), erecti (bg-en)
Noisy tokens นั้น ฉ่งกุ๊ ห ่ K ่ K

Contributions

- ► We detect unargmaxable classes unambiguously.
- Our algorithm handles the Softmax bias term.
- ► We search for unargmaxable tokens in 150 models.

Combined algorithm for detecting unargmaxable classes.

Use algorithm to search released models

* Breaking the Softmax Bottleneck: A High-Rank RNN Language Model, Yang et al., 2018
[†] Stolen Probability: A Structural Weakness of Neural Language Models, Demeter et al., 2020

Code github.com/andreasgrv/unargmaxable and visualisation unargmaxable.ai.



Practitioners need not worry about unargmaxable tokens, but loss in expressivity may still be an important problem.





Engineering and Physical Sciences Research Council

