

# A Head-to-Head Comparison of Perceptron and Logistic Regression for Morphological Tagging

Saman Rahimi

Stockholm University

sara5578@student.su.se

June 5, 2025

## Abstract

This paper compares the Perceptron and Multinomial Logistic Regression (MLR) for morphological feature prediction on the Swedish UniMorph dataset, a resource rich in typological diversity. Both models use identical prefix/suffix and POS-based binary features and are trained for three epochs. The task is to predict morphological labels such as number, definiteness, and gender, which are crucial for downstream NLP applications such as parsing and machine translation. Evaluation is conducted using accuracy as the main metric, with a paired bootstrap test applied to assess statistical significance. Although MLR is theoretically more powerful as a probabilistic model, the Perceptron slightly outperforms it (66.2% vs. 65.8%), though the difference is not statistically significant. These findings emphasize the robustness of mistake-driven learning in settings characterized by sparse features and highly diverse label spaces.

## 1 Background

Morphological feature classification involves assigning labels that capture grammatical properties such as number (e.g., singular/plural), definiteness, gender, or tense. In richly inflected languages like Swedish, this is a non-trivial task due to the large number of possible tag combinations and the sparsity of some forms in training data.

Two commonly used approaches for classification are the Perceptron and Multinomial Logistic Regression (MLR). The Perceptron is a mistake-driven algorithm that updates weights only when a prediction is incorrect. In contrast, MLR is a probabilistic model that updates weights based on the gradient of a log-likelihood function using all class probabilities (1). Both methods can be trained on

binary feature representations derived from POS, prefixes, and suffixes.

Accurate morphological classification is essential for downstream tasks such as dependency parsing and machine translation, especially in morphologically complex languages. Our study compares these two algorithms under identical training conditions using accuracy as the evaluation metric, with statistical significance assessed via the paired bootstrap method.

## 2 Methodology

We use the Swedish UniMorph dataset, where each line contains a wordform, its POS tag, and a set of morphological features. Our task is to predict the morphological features (e.g., SG;DEF) given a word and its POS.

**Feature Extraction:** For both models, we extract binary features consisting of the part-of-speech (POS), and all prefixes and suffixes of the word up to length five. Each feature is combined with the POS to increase disambiguation capacity. For example, for the word *boken* and POS *NOUN*, the features include NOUN prefix=b, NOUN suffix=en, etc.

**Labels:** The label to be predicted is the morphological tag (e.g., SG;DEF), excluding the POS tag.

### Models:

**Perceptron:** A standard online mistake-driven learner. On each error, it updates the weights by  $\pm 1$  for the predicted and true classes.

**Multinomial Logistic Regression (MLR):** A probabilistic classifier that computes softmax scores for all classes and updates weights using the gradient of the log-likelihood function.

**Training Setup:** Both models are trained for 3 epochs using the same training and test splits, fea-

tures, and label sets. Learning rate is set to 1.0 for both models, and no shuffling or regularization is applied.

**Evaluation:** We compute classification accuracy on the test set. To compare the two models, we use the paired bootstrap test (Jurafsky & Martin, 2024, §4.9) with 5000 resampling iterations to determine if differences are statistically significant. tag (e.g., SG;DEF) given the word and its POS.

**Features:** Binary features are extracted from all prefixes and suffixes (up to length five), combined with POS.

**Labels:** Only morphological tags, excluding the POS, are used as labels.

**Models:**

- **Perceptron:** Mistake-driven learner with  $\pm 1$  weight updates.
- **MLR:** Probabilistic classifier using softmax and gradient-based updates.

Both models are trained for 3 epochs with learning rate 1.0 on the same data and features. Accuracy is used for evaluation, and the paired bootstrap test is applied to assess significance.

### 3 Experiments and Results

Both models were trained and evaluated on the Swedish UniMorph dataset using identical training and test splits. Feature extraction and label encoding were kept consistent, using POS + prefix/suffix features and morphological tags excluding POS. Both models were trained for 3 epochs with a learning rate of 1.0.

Model	Accuracy (%)	Epochs
Perceptron	66.24	3
MLR	65.80	3

Table 1: Accuracy comparison between Perceptron and MLR on Swedish UniMorph.

**Observation:** While Perceptron achieved marginally higher accuracy, the difference between the two models was minimal.

**Statistical Significance:** We applied a paired bootstrap test with 5000 resampling iterations to assess the statistical reliability of the observed accuracy difference. The 95% confidence interval for

the difference in accuracy was:  $[-0.0068, 0.0156]$ . Since this interval includes zero, the difference is not statistically significant.

**Error Trends:** Both models showed reduced performance on rare morphological tags, especially low-frequency combinations of definiteness and plurality. Incorporating additional features such as character n-grams or subword representations may help improve generalization.

### 4 Discussion and Conclusion

Our results indicate that, despite the theoretical strengths of Multinomial Logistic Regression (MLR), the Perceptron achieves slightly higher accuracy in morphological feature classification on the Swedish UniMorph dataset. However, the paired bootstrap test shows that this difference is not statistically significant, suggesting that both models perform comparably in practice.

The strong performance of the Perceptron may be attributed to its robustness in handling sparse, binary feature spaces, which aligns well with our prefix/suffix-based features. Moreover, since the model uses direct mistake-driven updates and avoids computing full probability distributions, it may be better suited for limited data settings where overfitting is a concern. The binary nature of the features and the hard decision boundaries of Perceptron likely contribute to its success in this parsing-oriented task.

In contrast, MLR, while probabilistically grounded, may be more sensitive to feature imbalance and lacks the strong bias toward conservative updates seen in Perceptron. This may lead to suboptimal generalization unless combined with regularization techniques.

This result aligns with previous observations in NLP, where simpler online algorithms outperform more complex probabilistic models under constrained or sparse feature settings.

Future work could involve evaluating these models on other morphologically rich languages, incorporating richer features such as character n-grams or POS embeddings, and applying regularization strategies in MLR to improve robustness.

## References

- [1] Daniel Jurafsky and James H. Martin. 2024. *Speech and Language Processing* (3rd ed. draft). Sections 4.9, 5.3, 5.8. <https://web.stanford.edu/~jurafsky/slp3/>
- [2] Ryan Cotterell, Christo Kirov, John Sylak-Glassman, David Yarowsky, Jason Eisner, and Mans Hulden. 2018. The UniMorph 2.0 Project: Universal Morphological Inflection Across Languages. In *Proceedings of LREC 2018*.
- [3] Michael Collins. 2002. Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In *EMNLP 2002*.
- [4] Yoav Goldberg. 2017. *Neural Network Methods for Natural Language Processing*. Morgan & Claypool Publishers.
- [5] Mark Dredze, Koby Crammer, and Fernando Pereira. 2008. Confidence-weighted linear classification. In *ICML 2008*.
- [6] Kyunghyun Cho et al. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *EMNLP 2014*.
- [7] Bradley Efron and Robert Tibshirani. 1994. *An Introduction to the Bootstrap*. Chapman & Hall.
- [8] Saman Rahimi. 2026. Code for Perceptron and Multinomial Logistic Regression used in morphological classification experiments. Available at: <https://github.com/samanrh/morph-feature-classifier.git>