CIS5300 - Speech and Language Processing - Chapter 5 Notes

Jonathon Delemos - Chris Callison Burch

June 29, 2025

0.1 Abstract - Logistic Regression

In this chapter we introduce an algorithm that is admirably suited for discovering the link between features or clues and some particular outcome: **logistic regression**. Indeed, logistic regression is one of the most important analytic tools in the social and natural sciences. In natural language processing, logistic regression is the base- line supervised machine learning algorithm for classification, and also has a very close relationship with neural networks. As we will see in Chapter 7, a neural network can be viewed as a series of logistic regression classifiers stacked on top of each other. Thus the classification and machine learning techniques introduced here will play an important role throughout the book

0.2 *Components of Probabilistic Machine Learning Classifier*

- **Feature Representation** - for each input, there will be vector of features.
- **Sigmoid and Softmax** - tools for classification. We will show the formulas needed later.
- **Cross-Entropy Loss Function** - minimizing the loss corresponding to error in training.
- **Stochastic Gradient Descent** - This means updating the weighted variables each time a vector is processed, as opposed to waiting for a batch transaction then updating the weights.

0.3 The Sigmoid Classifier/Function

For this step of our probabalistic machine learning classifier, we will focus on making a simple value that is a sum of the weights multiplied by the input parameter added with the bias term. This is the **sigmoid classifier (z)**. The bias (b) can be thought of as a base *y-intercept* for the model.

$$z = (\sum_{i=1}^{n} w_i x_i) + b$$

This sigmoid value can range anywhere between - ∞ and ∞. It is not bound by the legal probabilties of 0-1, which we will see later.
To create a probability, we'll pass the z through the **sigmoid function** $\sigma(z)$. This is also called the logistic function.

$$\sigma(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+exp(-z)}$$

We are almost there. If we apply the sigmoid to the sum of the weighted features, we get a number 0 and 1. To make it a probability, we just need to make sure that the two cases p(y=1) and p(y=0) sum to 1.

$$P(y=1) = \sigma(w \cdot x + b)$$

Recall:

$$\sigma(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+exp(-z)}$$

Substitution for the z:

$$= \frac{1}{1 + exp(-(w \cdot x + b))}$$

Next, the probability that y = 0.

$$P(y = 0) = \frac{exp(-(w \cdot x + b))}{1 + exp(-(w \cdot x + b))}$$

0.4 5.2 - Classification with Logistic Regression