(https://google.com/racialequity?authuser=0)

Multi-Class Neural Networks: Softmax

ated Time: 8 minutes

Recall that logistic regression (https://developers.google.com/machine-learning/crash-course/logistic-regression?authuser=0) produces a decimal between 0 and 1.0. For example, a logistic regression output of 0.8 from an email classifier suggests an 80% chance of an email being spam and a 20% chance of it being not spam. Clearly, the sum of the probabilities of an email being either spam or not spam is 1.0.

Softmax extends this idea into a multi-class world. That is, Softmax assigns decimal probabilities to each class in a multi-class problem. Those decimal probabilities must add up to 1.0. This additional constraint helps training converge more quickly than it otherwise would.

For example, returning to the image analysis we saw in Figure 1, Softmax might produce the following likelihoods of an image belonging to a particular class:

Class	Probability
apple	0.001
bear	0.04
candy	0.008
dog	0.95
egg	0.001

Softmax is implemented through a neural network layer just before the output layer. The Softmax layer must have the same number of nodes as the output layer.

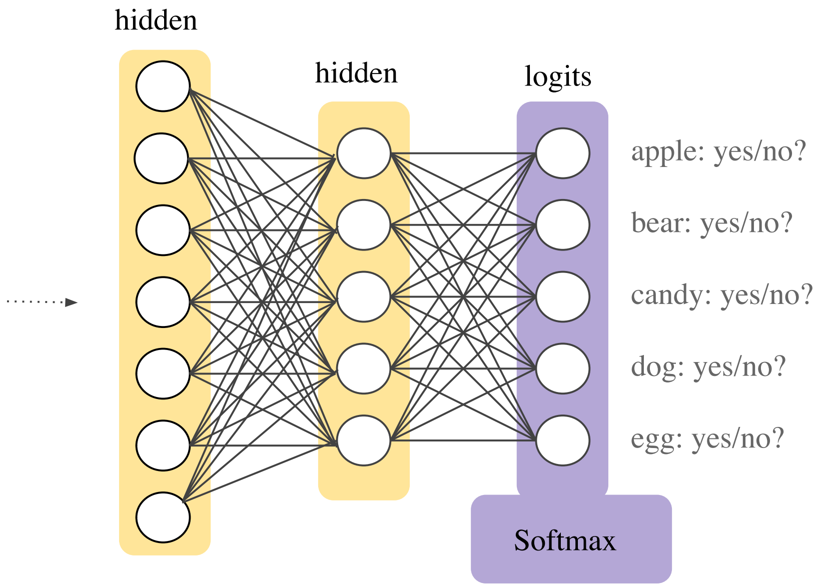

Figure 2. A Softmax layer within a neural network.

+ **Click the plus icon to see the Softmax equation.**

The Softmax equation is as follows:

$$p(y = j|\mathbf{x}) = \frac{e^{(\mathbf{w}_j^T\mathbf{x}+b_j)}}{\sum_{k\in K} e^{(\mathbf{w}_k^T\mathbf{x}+b_k)}}$$

Note that this formula basically extends the formula for logistic regression into multiple classes.

Softmax Options

Consider the following variants of Softmax:

- **Full Softmax** is the Softmax we've been discussing; that is, Softmax calculates a probability for every possible class.
- **Candidate sampling** means that Softmax calculates a probability for all the positive labels but only for a random sample of negative labels. For example, if we are interested in determining whether an input image is a beagle or a bloodhound, we don't have to provide probabilities for every non-doggy example.

Full Softmax is fairly cheap when the number of classes is small but becomes prohibitively expensive when the number of classes climbs. Candidate sampling can improve efficiency in problems having a large number of classes.

One Label vs. Many Labels

Softmax assumes that each example is a member of exactly one class. Some examples, however, can simultaneously be a member of multiple classes. For such examples:

- You may not use Softmax.
- You must rely on multiple logistic regressions.

For example, suppose your examples are images containing exactly one item—a piece of fruit. Softmax can determine the likelihood of that one item being a pear, an orange, an apple, and so on. If your examples are images containing all sorts of things—bowls of different kinds of fruit—then you'll have to use multiple logistic regressions instead.

rms

- ndidate sampling ://developers.google.com/machine-ıg/glossary?authuser=0#candidate_sampling)
- ılti-class ://developers.google.com/machine-ıg/glossary?authuser=0#multi-class)
- logistic regression (https://developers.google.com/machine-learning/glossary?authuser=0#logistic_regression)
- softmax (https://developers.google.com/machine-learning/glossary?authuser=0#softmax)

Help Center (https://support.google.com/machinelearningeducation?authuser=0)

Previous
← One vs. All (https://developers.google.com/machine-learning/crash-course/multi-class-neural-networks/one-vs-all?authuser=0)

Next
Programming Exercise → (https://developers.google.com/machine-learning/crash-course/multi-class-neural-networks/programming-exercise?authuser=0)