**Cross-entropy Loss and Softmax Classifiers**

Softmax classifiers are very popular in the context of deep learning and convolutional neural networks and are used in this study (section 3.6.1). Softmax classifiers give the probabilities for each class label, where it is much easier for humans to interpret probabilities, rather than margin scores as it is usually done in Multi-class SVM Loss. In cross-entropy loss, the mapping function *f* is defined such that it takes an input set of data *xi* and maps them to output class labels via dot product of the data *xi* and weight matric *W*:

(1)

The scoring function is defined as:

(2)

which implies that we can obtain the predicted score of the *jth* class via the *ith* data point:

(3)

The scores can then be interpreted as unnormalised log probabilities for each class label, which mounts to swapping out the hinge loss function with cross-entropy loss:

(4)

The loss function should minimise the negative log likelihood of the correct class:

(5)

The probability statement can be interpreted as:

(6)

Where we use the standard scoring function form:

(7)

This yields the final loss function for a single data point, as above:

(8)

It should be noted that the algorithm here is actual base *e* (natural logarithm) since we are taking the inverse of the exponentiation over *e*. The actual exponentiation and normalisation via the sum of exponents is the *Softmax function*. The negative log yields the actual cross-entropy loss.

Finally, computing the cross-entropy loss over an entire dataset is done by taking the average:

(9)

We refer readers to(Marsland 2009; Harrington 2012; Rosebrock 2019) for more details about the concept of loss functions in machine learning.

Harrington P. Machine Learning in Action. 2012;5:384.   
  
Marsland S. Machine Learning: An Algorithmic Perspective. 2009;   
  
Rosebrock A. Deep Learning for Computer Vision. 2019;