Multiclass Prediction

**SoftMax Regression, One-vs-All & One-vs-One for Multi-class Classification**

In Multi-class classification, we classify data into multiple class labels. Unlike classification trees and nearest neighbors, the concept of Multi-class classification for linear classifiers is not as straightforward. We can convert logistic regression to Multi-class classification using multinomial logistic regression or SoftMax regression; this is a generalization of logistic regression. SoftMax regression will not work for Support Vector Machines (SVM); One vs. All (One-vs-Rest) and One vs One are two other multi-class classification techniques that can convert most two-class classifiers to a multi-class classifier.

# SoftMax Regression

SoftMax regression is similar to logistic regression, the SoftMax function converts the actual distances i.e. dot products of 𝑥 with each of the parameters ​ for 𝐾 classes in the range from 0 to 𝐾-1. This is converted to probabilities using the following formula.

The training procedure is almost identical to logistic regression using cross-entropy, but the prediction is different. Consider the three-class example where y i.e y can equal 0,1,2. We would like to classify x. We can use the SoftMax function to generate a probability of how likely the sample belongs to each class. We then make a prediction using the function:

Let’s do an example, consider sample we will start by creating a table where each column will be the values of the SoftMax function. The index of each column is the same as the class.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

**Table 1. Each column will be the i-th values of the SoftMax function. The index of each column is the same as the class.**

Let’s add some real probabilities, this is the models estimate of how likely a sample belongs to each class.

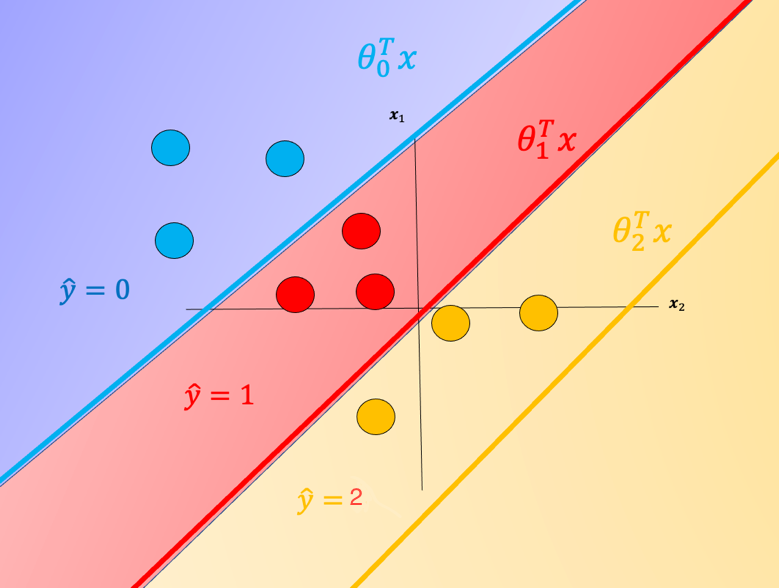
|  |  |  |
| --- | --- | --- |
| **0.97** | **0.02** | **0.01** |
|  |  |  |

**Table 2. Table of real probabilities. Each column will be the i-th values of the SoftMax function. The index of each column Is the same as the class.**

We can represent the probability as a vector [0.97,0,02,0.01]. To get the class we simply apply the *argmax* function, this returns the index of the largest value.

**Geometric Interpretation**

Each  is the equation of a hyperplane, we plot the intersection of the three hyperplanes with 0 in fig 1 as colored lines, in addition, we can overlay several training samples. We also shade the regions where the value of is largest, this algo corresponds to the largest probability. This color corresponds to where a sample x would be classified. For example if the input is in the blue region, the sample would be classified , if the input is in the red region it would be classified as , and in the yellow region . We will use this convention going forward.



#### **Fig 1. Equation of a hyperplane. We plot the intersection of the three hyperplanes with 0, in addition we can overlay several samples.  We also shade the regions where the value of i is largest.**

One problem with SoftMax regression with cross-entropy is it cannot be used for SVM and other types of two-class classifiers.