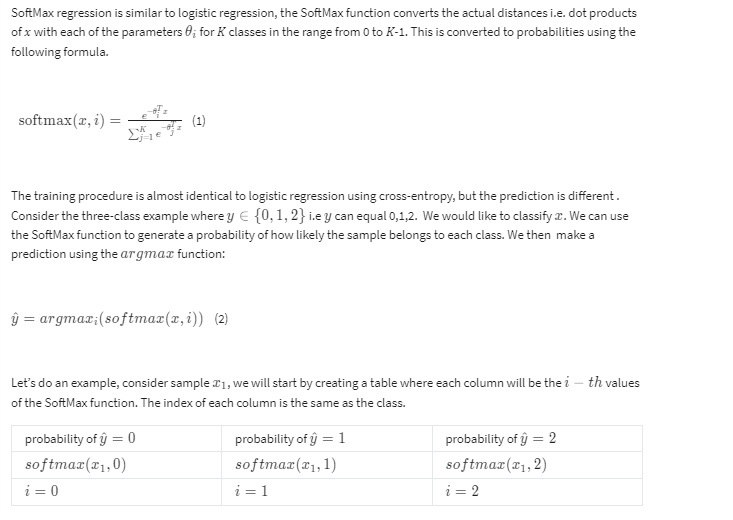
**Multiclass Prediction**

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

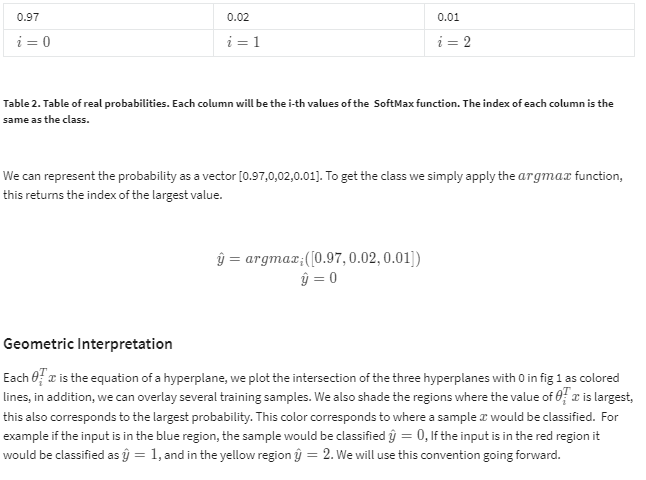
In Multi-class classification, we classify data into multiple class labels. Unlike classification trees and nearest neighbors, the concept of Multi-class classification for linear classifiers is not as straightforward. We can convert logistic regression to Multi-class classification using multinomial 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.

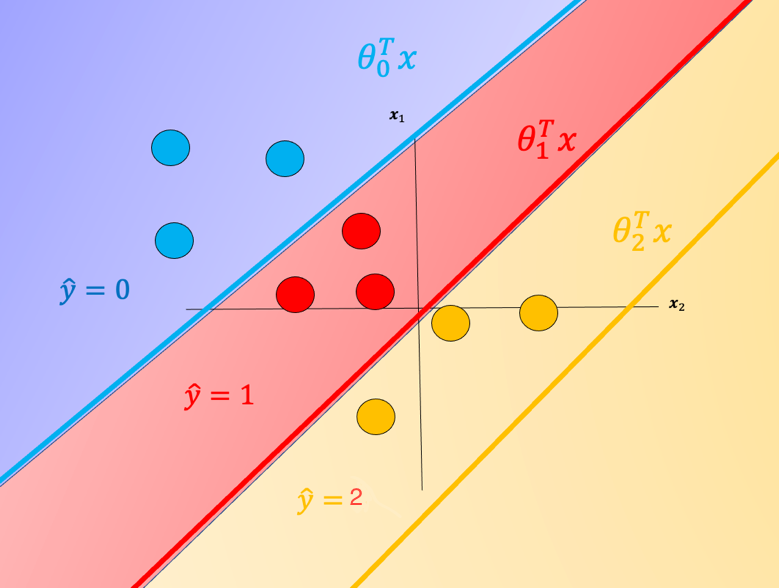
**SoftMax Regression**



**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.

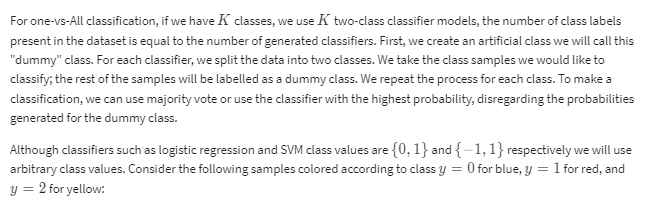
****

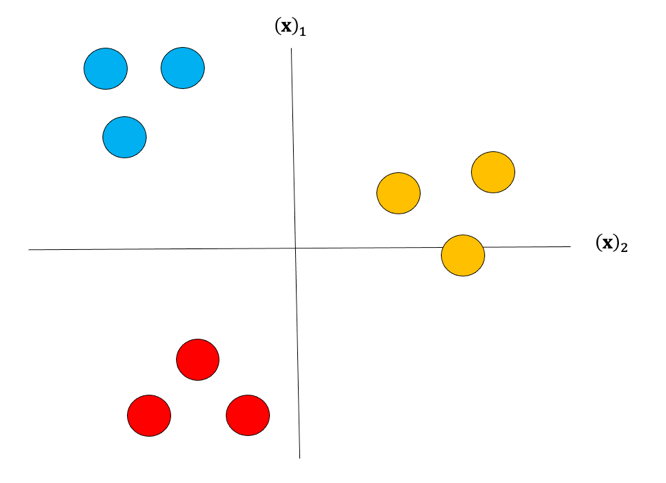


**Fig 1. *E*quation of a hyperplane. We plot the intersection of the three hyperplanes  with 0, in addition  we can overlay several samples.  We also shade the regions where the value of *i* is largest.**

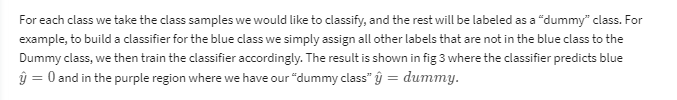
One problem with SoftMax regression with cross-entropy is it cannot be used for SVM and other types of two-class classifiers.

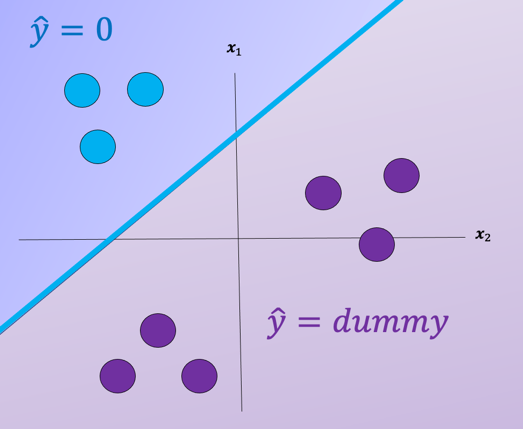
**One vs. All (One-vs-Rest)**

****



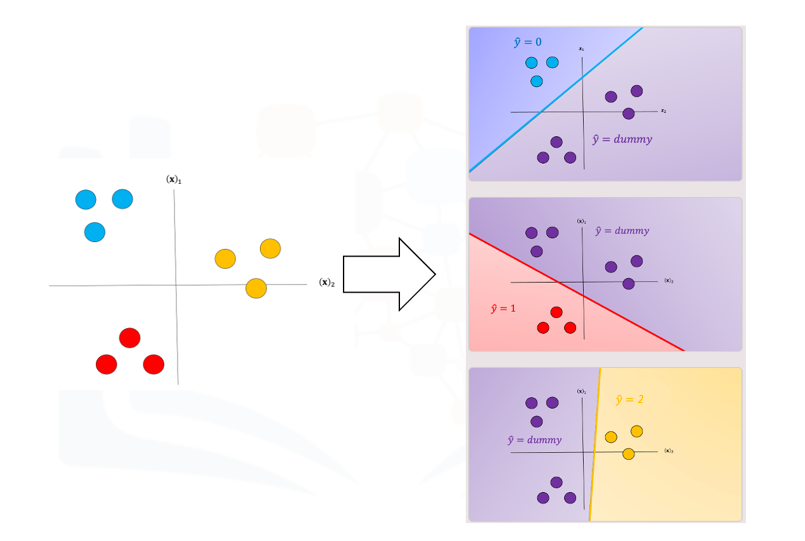
**Fig 2. Samples colored according to class.**

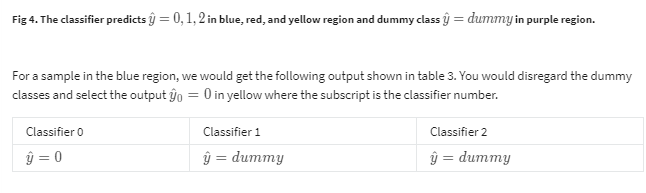
****



****

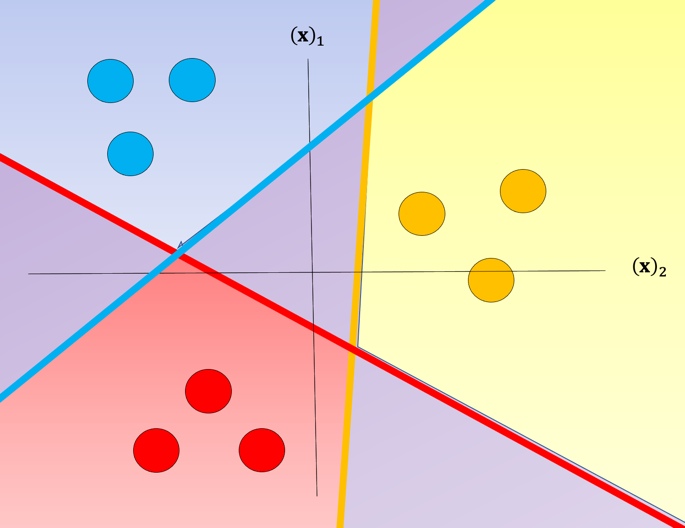
We repeat the process for each class as shown in Fig 4, the actual class is shown with the same color and the corresponding dummy class is shown in purple. The color of the space is the actual classifier predictions shown in the same manner as above.

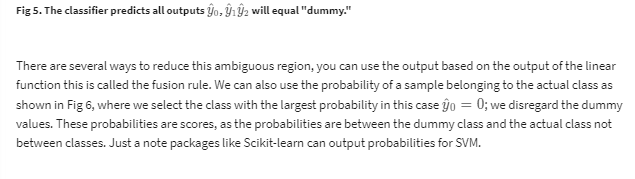


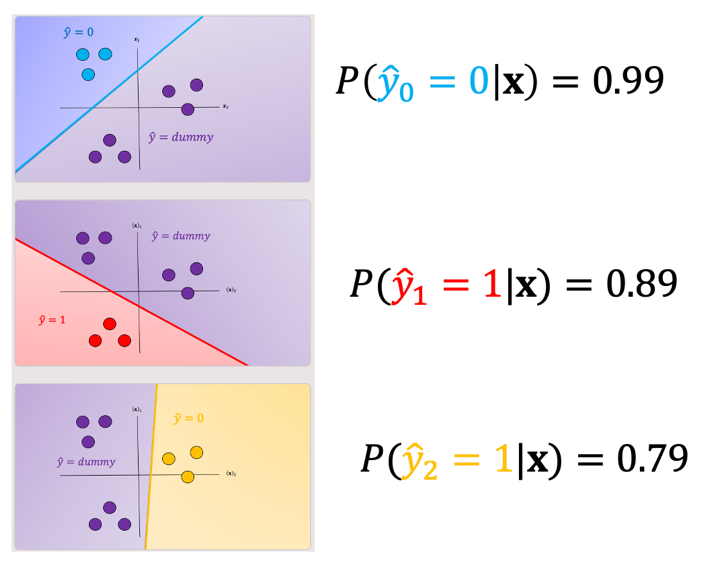
****

**Table 3. Example classification output, 2 of the 3 outputs are dummy; these classifiers would be ignored and the class would be zero.**





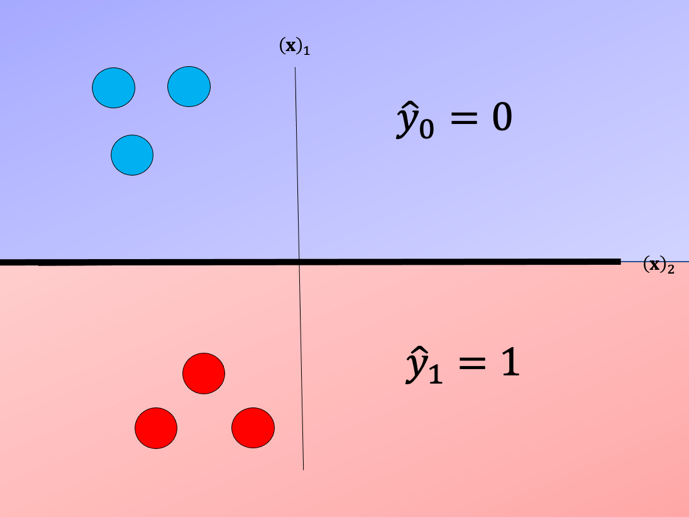


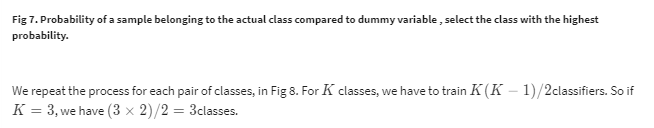


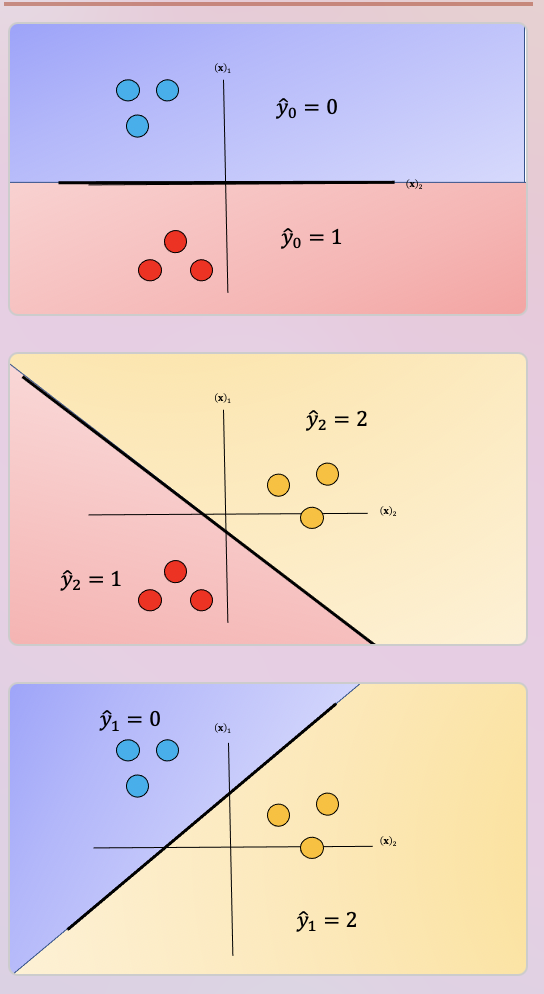
**Fig 6. Probability of a sample belonging to the actual class compared to dummy variable, selects the class with the highest probability.**

**One-vs-One classification**

In One-vs-One classification, we split up the data into each class; we then train a two-class classifier on each pair of classes. For example, if we have class 0,1, and 2, we would train one classifier on the samples that are class 0 and class 1, a second classifier on samples that are of class 0 and class 2, and a final classifier on samples of class 1 and class 2. Fig 7 is an example of class 0 vs class 1, where we drop training samples  of class 2.  Using the same convention as above where the color of the training samples are based on their class. The separating plane of the classifier is in black.  The color represents the output of the classifier for that particular point in space.







**Fig 8.  Probability of a sample belonging to the actual class compared to dummy variable, select the class with the highest probability.**

To perform Classification on a sample, we perform a majority vote where we select the class with the most predictions.  This is shown in Fig  9 where the black point represents a new sample and the output of each classifier is shown in the table. In this case, the final output is one as selected by two of the three classifiers. There is also an ambiguous region but it’s smaller, we can also use similar schemes as in One vs all like the fusion rule or using the probability. Check out the labs for more.

