First I’ll talk about the nature of our problem, and then the methods to solve nonlinear multiclass classification.

Since our data set has a lot of attributes and due to the nature of football games, it is a nonlinearly separable dataset. There are three classes in our data sets, namely, win, draw and loss. Because both Chelsea and Arsenal are successful members of the premier league, they generally won their matches against Yo-Yo clubs. Therefore, while collecting the data from internet sites, we have observed that our data set is imbalanced. Apart from nonlinearity and multiclass classifications, this result has added more complexity to our program. **(next slide\*)**

Having covered the nature of our problem, now let me move on to the second topic, methods to solve nonlinear multiclass classifications.

Because we have implemented our project with SVM’s I’ll start with methods on SVM’s. There are two methods, 1 is based on binary classification and the other one is based on all-together methods.

In binary classification, there are two methods, 1 against all and one against one. In the first one, k SVM models are constructed, and k denotes here the number of classes. The i’th SVM is labeled with positive labels and all other classes are labeled with negative labels. In the second one, the one and against one method, classifiers are constructed where each one is trained on the data from two classes.

In the all-together methods, results are obtained by solving one problem. Either we can do a similar approach to the one versus all or we can solve this problem by single optimization problem as it is proposed by Crammer and Singer. **(next slide\*)**

**Another candidate method** for our problem was Decision Tree learning. Because our problem has a high number of dimensions, it is really hard to visualize it and because of ease of application, we didn’t choose decision tree learning. Also, we were not sure that whether the decision tree learning package on the internet is sufficient for our problem or not and we find that tuning the algorithm is difficult for us. Moreover, in some research, it is said that the decision tree algorithm might not be suitable for the imbalanced data sets. It may overfit the data. (next slide\*)

**The third candidate method** for our problem was Artificial Neural Networks aka ANN’s. It is currently one of the most widely used models in Medline. However, ANN mainly depends on the quality of the dataset, the care with the model parameters and the evaluation criteria. Therefore, it offers a great flexibility between input and output neurons, however, it carries a high risk of overfitting.

Also another difficulty in deep neural networks using gradient descent. To train the hidden layers, very large number of epochs may be needed because a neuron with sigmoid activation function may become saturated. And its derivation becomes very small. To solve this there are several methods, but I’m not going into the details. Other than saturation, the initialization of the weights can have a different impact on the model. It is hard for us to do well-initialized networks.