# Question #1

An experimental team from the Institute of Agrophysics in Poland was interested in the ability of a soft X-ray technique to differentiate different types of wheat seeds. The interest lay in the fact that they wished to study what had been planted in historic wheat fields, and more high-tech means of determining the type of wheat were outside of their budget. For this experiment, the group calculated various properties of the wheat seeds using the X-ray technique. Their data can be found at

<https://archive.ics.uci.edu/ml/datasets/seeds>

Please carefully study the data, as there may be some missing values, marked as “N/A”. You will need to replace these with the mean of the particular feature in which the N/A occurs. Please use a neural network to classify the seeds. Play with parameters, such as number of nodes in a hidden layer and number of hidden layers. You will find the pseudo-code titled “Classification with neural nets” in the Resources/Code folder on Collab useful for this endeavor. Do you think the X-ray technique was adequate for deciphering amongst the seeds?

**Extra Credit**: Use a cross-validation approach to train and test the data.

# Question #2 – Support Vector Machines

1. Using the S&P 500 data posted on UVAcollab, manually fit the support vector classifier (i.e. set up and run the optimization problem) to forecast the final day of the series (12/31/2014). You can either set up the quadratic program directly in R or import the data into MS Excel and run the optimization using the Solver.
   1. **Target Variable:** Binary classification for S&P 500 Return Positive (Rt ≥0) or Negative (Rt <0)
   2. **Predictors:** Rt-1 & Rt-2
   3. **Training Window:** Trailing 126 trading days (~6 months)
2. Plot the training data as well as the decision boundary and the margins. Identify the support vectors.
3. Formulate and solve the dual problem. Do the solution and support vectors match those found in a)?
4. Suppose you wanted to add a third “zero-return” class for small returns less than a specified threshold µ in absolute value. Repeat a) and b) for both the OVA and OVO methods of extending SVMs to multi-class problems with µ≡0.1%.
   1. **Target Variable:** Classification Return Positive (Rt ≥µ≡0.1%), “Zero” (µ ≥ Rt ≥-µ≡0.1%), or Negative (Rt <-µ)
   2. **Predictors:** Rt-1 & Rt-2
   3. **Training Window:** Trailing 126 trading days (~6 months)
5. Comment on the overall results.

# Question #3 – Tree-based Methods

1. Using the S&P 500 data posted on UVAcollab, manually fit a full classification tree on for the problem in 2.a) using misclassification rate as the minimization criteria and stopping when all nodes have less than 12 training points. Again, you may set up the optimization directly in R or import the data into MS Excel and run the successive optimizations using the Solver.
   1. **Target Variable:** S&P 500 Return Positive (Rt >=0)
   2. **Predictors:** Rt-1 & Rt-2
   3. **Training Window:** Trailing 126 trading days (~6 months)
2. Prune the tree based on deviance to obtain the best 5 nodes and draw the resulting tree.
3. Plot the training data and the decision boundaries from the tree in b).