

Q.1 Question 2.4 from textbook (Tom Mitchell).

Q.2 **Explain** two drawbacks each of Find-S and Candidate Elimination algorithms. Develop a new algorithm for concept learning, which overcomes at least one of the major drawbacks listed above.

Q3. **Explain** supervised, unsupervised and semi-supervised learning paradigms. Give real life examples for each.

Q4. **Programming Question:** Download the Skin Segmentation Dataset

(<http://archive.ics.uci.edu/ml/datasets/Skin+Segmentation>) for a two class problem. Perform 3-fold cross validation on the dataset with the following algorithm (create folds such that each fold has equal number of samples from both classes):

1. Calculate the range, mean and variance of each attribute (using the training set only). Which attribute seems most 'consistent' with respect to the two classes?
2. Create a histogram for each attribute for the two classes. You should have three histograms, one for each attribute.
3. Analyze the histograms and select the attribute which appears to be having the most discriminatory behavior for the given problem. (You can have a completely different answer from your friend. You should just be able to justify it.)
4. Using the attribute selected in the previous step and simple thresholding, perform classification for the test set. Report the true positive, false positive, true negative and false negative rates.
5. Pick two attributes at a time and create a scatter plot with the training data, for the two classes. You should have three such plots.
6. Based on the plots above, which pair appears to have the most discriminatory behavior for the given two class problem?

Q5. Perform any two data normalization techniques on the data provided (remember, for any technique, you would only have the training data for selecting normalization parameters). Repeat Q4 on the normalized data for any one fold. Analyze the effect of normalization on this dataset.

Q6. Find two Machine Learning researchers/scientists who contributed significantly to the field. In the spreadsheet shared yesterday evening ("ML Presentation Schedule"), in the second worksheet, list the names along with their major contribution (not more than one sentence). **Bonus:** Find an individual who contributed significantly who was **not** of the Machine Learning domain. **Caution:** Before filling up a name, make sure it has **NOT** been mentioned before.