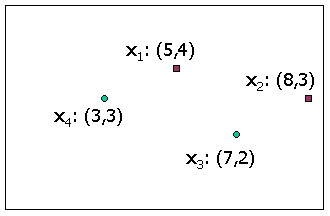
### **Machine Learning**

The first two questions involve the following dataset, which has two positive points (purple squares) and two negative points (green circles).



That is, the training data set consists of:

(**x**1,y1) = ((5,4),+1)  
(**x**2,y2) = ((8,3),+1)  
(**x**3,y3) = ((7,2),-1)  
(**x**4,y4) = ((3,3),-1)

**Question 1**: This data set is separable. If we call the horizontal axis of the space *u* and the vertical axis *v*, then a form of the decision boundary is *v*=*c*+*au*. Note that in this form, *c* is the intersection of the boundary with the vertical (*v*) axis, and *a* is the slope.

In all cases, if (u,v) is one of the points **x**1 or **x**2, then the point will be above the boundary (that is, if **x**1 = (u,v), then *v* >= *c*+*au*, and similarly for **x**2). Likewise, **x**3 and **x**4 must be below the boundary.

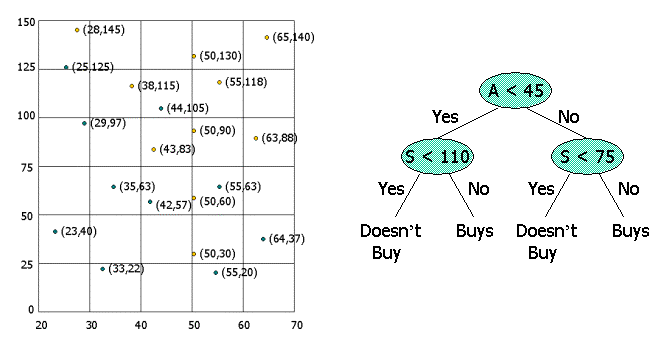
Various values of *a* and *c* could be chosen, but there are significant constraints on what *a* and *c* can be. Deduce those constraints, and then find a possible value of *a* and *c*.

**Question 2**: Our goal is to find the maximum-margin linear classifier for this data. In easy cases, the shortest line between a positive and negative point has a perpendicular bisector that separates the points. If so, the perpendicular bisector is surely the maximum-margin separator. Alas, in this case, the closest pair of positive and negative points, **x**2 and **x**3, have a perpendicular bisector that misclassifies **x**1 as negative, so that won't work.

The next-best possibility is that we can find a pair of points on one side (i.e., either two positive or two negative points) such that a line parallel to the line through these points is the maximum-margin separator. In these cases, the limit to how far from the two points the parallel line can get is determined by the closest (to the line between the two points) of the points on the other side. For our simple data set, this situation holds.

Consider all possibilities for boundaries of this type, and express the boundary as **w**.**x**+*b*=0, such that **w**.**x**+*b*>=1 for positive points **x** and **w**.**x**+*b*<=-1 for negative points **x**. Assuming that **w** = (*w*1,*w*2), list out the values of *w*1, *w*2, and *b*.

**Question 3**: Below we see a set of 20 points and a decision tree for classifying the points.

 To be precise, the 20 points represent (Age,Salary) pairs of people who do or do not buy gold jewelry. Age (abbreviated A in the decision tree) is the x-axis, and Salary (S in the tree) is the y-axis. Those that do are represented by gold points, and those that do not by green points. The 10 points of gold-jewelry buyers are:

(28,145), (38,115), (43,83), (50,130), (50,90), (50,60), (50,30), (55,118), (63,88), and (65,140).

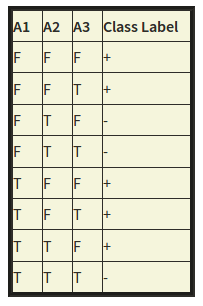
The 10 points of those that do not buy gold jewelry are:

(23,40), (25,125), (29,97), (33,22), (35,63), (42,57), (44, 105), (55,63), (55,20), and (64,37).

Some of these points are correctly classified by the decision tree and some are not. Determine the classification of each point, and then indicate the points that are misclassified.

**Question 4**: Consider the process of building a binary classifier based on a decision-tree model using the Gini Index as a measure of impurity associated with a tree node that represents a subset of training examples. A node is split into partitions represented by its child nodes based on the values of a selected attribute. The goodness of the attribute for the split, referred to as gain of the attribute, is estimated in terms of the difference between the impurity of the parent node and the weighted sum of the impurities of the child nodes.

Consider the training set of examples described in terms of three attributes A1, A2 and A3 in addition to the decision attribute, Class Label.



The first three columns of the table represent the values of the three binary attributes and the last column indicates whether the element (described by the row) is a positive example of the class or not (+ indicates a positive example). The classifier aims at classifying the training examples into positive or negative examples of the class under consideration.

Estimate the goodness of the attributes for splitting the training set given in the table and explain why?

**Question 5**: Suppose our training set consists of the three negative points (1,4), (3,3), and (3,1) and the two positive points (3,6) and (5,3). If we use nearest-neighbor learning, where we classify a point to be in the class of the nearest member of the training set, what is the boundary between the positive and negative points? Identify the points that are classified as positive.