$\qquad$ KEY $\qquad$
$\qquad$

1) We have learned about four machine learning algorithms, including the decision tree.

If I had to describe the decision tree in one or two sentences, I might say:
A decision tree classifies an unknown instance by asking a series of yes/no questions. These yes/no questions are represented as interval nodes in a tree, and leaf nodes contain the predicted class labels; thus, a classification of an unknown instance can be seen as a traversal from the root node to a leaf node, based on the answers to the questions.

Pick any two of the other four algorithms, and explain them at the same level of detail.

The bayes classifier classifies an unknown instance by computing the probability it belongs to each class, and choosing the most probable class.

The nearest neighbor classifier classifies an unknown instance by finding its nearest neighbor (under some distance measure) and assigning it that item's label.

The linear-classifier classifies an unknown instance, by precomputing the line that best separates the two (or more) classes, and testing to see which side of the line the unknown instance is on.
2) Imagine we build a dendrogram, for five animals...
A. African elephant
B. Mouse
C. Indian elephant
D. Rat
E. House fly
...based on their size. Draw the dendrogram you might expect to see. Just put the letters A to E at the left nodes. Below I show a sample dendrogram to get you thinking.

$\qquad$ KEY $\qquad$
$\qquad$

|  | 3) Consider the problem to the left. Draw a decision tree that would work well for this problem. Clearly label the internal nodes with the question that will make the split, and label the outgoing branches with " Y " and " N " |
| :---: | :---: |
|  | $10$ |
| 4) What is the default rate here? $20 /(20+10)$ | $\begin{aligned} & 9 \\ & 8 \\ & 7 \end{aligned}$ |
| 5) What is the approximate accuracy of the simple linear classifier here? $20 / 30$ |  |
| 6) What is the approximate accuracy of the nearest neighbor classifier here? <br> About 29 or 30 out of 30 , or $95 \%$ to $100 \%$ |  |
| 7) What is the approximate accuracy of a decision tree classifier here? <br> About 29 or 30 out of 30 , or $95 \%$ to $100 \%$ | $\begin{array}{lllllllllll}1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10\end{array}$ |

8) Consider the problem below. The accuracy of my tree was $100 \%$. There was 20 datapoints that I use to learn this tree, 10 examples of sick ' O ', and ten examples of healthy ' X ', but I deleted the datapoints from the plot.
A) Draw in the 20 missing datapoints
B) Draw in the decision boundary's

