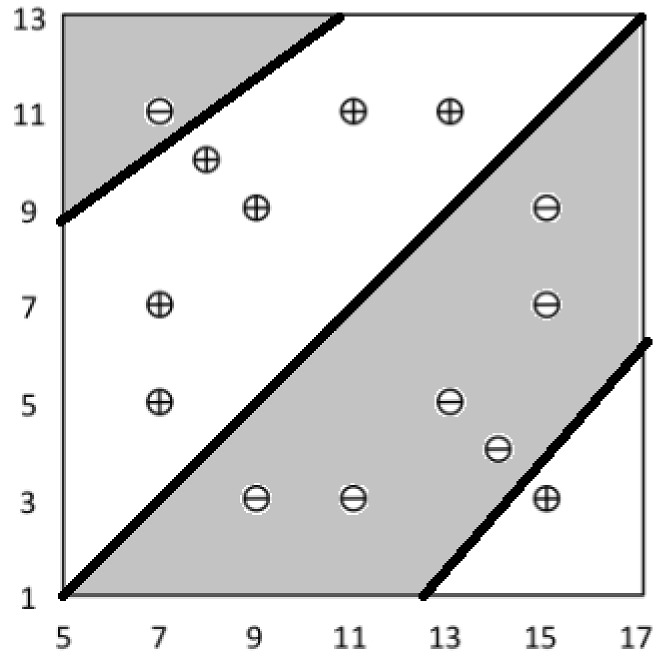
Logan Esch

1. Using leave-one-out cross-validation with a majority classifier is a bad idea when there are equal numbers of positive and negative examples. Suppose one of the positive examples is left out, then there are 99 positive examples and 100 negative. The majority classifier will pick negative for any input. The test case will be positive and thus the majority classifier will be wrong. The same is true if a negative example is left out for validation. This means that the majority classifier will be wrong all the time if there are equal numbers of positive and negative examples.



2. In the above figure, points falling in the shaded areas would be considered '-' while the non-shaded would be considered '+'.

   If the true concept is IF $x_2 > (x_1 - 4)$ THEN "+" ELSE "-" then we have not learned the concept exactly. For example, suppose we have a point (7, 12). Since 12 > (7-4) this point should be marked as a "+". However, using nearest neighbor with k=1 the nearest point to this one is d_1 which is a "-". Therefore we have not learned the concept exactly. Changing to k=3 would be able to mark (7, 12) correctly because the 2nd and 3rd closest points are "+" and since all three points are treated equally they "vote". Using k=14 does not make it better, this is for two reasons. First we are using all of our data, meaning we will categorically say that all new points are a member of whatever class is larger. In this case, the classes are tied and so we are left with nothing more than a random guess.

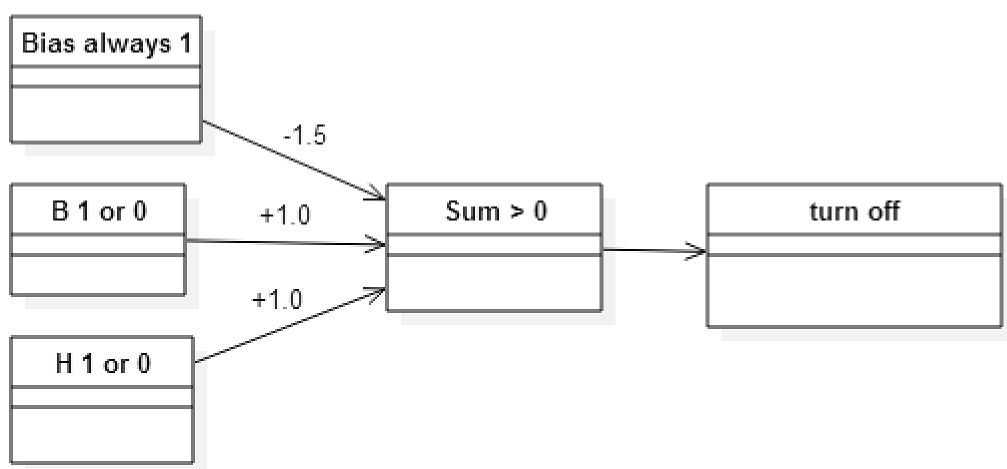
Logan Esch

For the next part of the problem, I am assuming that if there is a tie in distance for the third point being referenced between two points with opposing signs that k is reduced by one. If the remaining two points are not in agreement, then the closest point is chosen.

Training Set	Test Set	Number of Errors
2; : : : ; 14	1	1
1; 3; : : : ; 14	2	0
1; 2; 4; : : : ; 14	3	0
1; : : : ; 3; 5; : : : ; 14	4	0
1; : : : ; 4; 6; : : : ; 14	5	0
1; : : : ; 5; 7; : : : ; 14	6	0
1; : : : ; 6; 8; : : : ; 14	7	0
1; : : : ; 7; 9; : : : ; 14	8	0
1; : : : ; 8; 10; : : : ; 14	9	0
1; : : : ; 9; 11; : : : ; 14	10	0
1; : : : ; 10; 12; : : : ; 14	11	0
1; : : : ; 11; 13; 14	12	0
1; : : : ; 12; 14	13	0
1; : : : ; 13	14	1
Total Errors		2
Error Rate		1/7 = .143

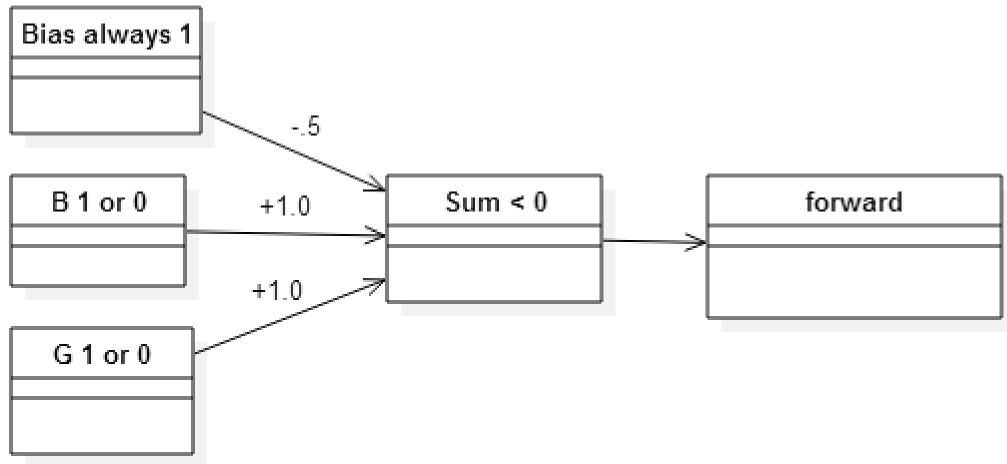
3. The agent shut off behavior can be described as:

   $B \wedge H \rightarrow$ off



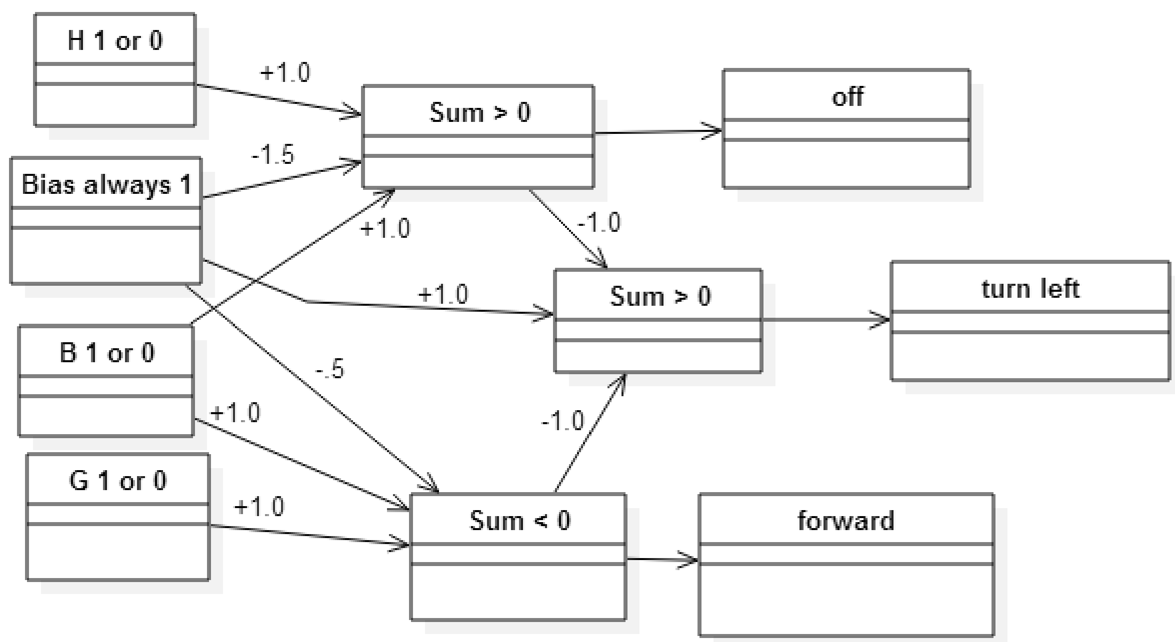
Logan Esch

¬B ∧ ¬ G → forward



(B ∧ ¬H) ∨ (¬B ∧ G) → turn-left

Because the function to describe turn-left is non-linear there cannot be a perceptron constructed that fires when and only when the agent should turn left.



4. For this problem I am assuming that there is at least one point in each class (positive, negative). If there is not then there is no need for any separator.
   Because all points are positive or negative there are two classes to which any point can belong.

Logan Esch

If there are n=3 points, then at least two of the points belong to the same class by the pigeon hole principle. This means that one point belongs to a different class. Any three points that are nonlinear form a triangle. Any line that bisects this triangle that does not contain one of the points must separate one point from the other two. Let the point that is separated be the point that belongs to the minority class, then there is a linear separator that will divide the two classes. If the points are all collinear they can still be divided linearly so long as the minority point is not between the majority points. If the minority point is between the majority points then any line separating the minority point from one of the majority points also separates the other majority point from the first.

If there are n=4 points in a 2d space then there is a linear separator between them if there is an uneven division between the classes (i.e. 3 positive, 1 negative or vice versa). Simply drop the farthest point from the minority class (as it contains no more information than the other two points) and now this problem is the above 3 points problem. If there are 2 points in each class then let there be line segments between them. If these line segments intersect, there cannot be a linear separator for the two classes. This is because any linear function that divides two points from the rest must divide two points that belong to opposing classes from the rest.

If there are n=4 points in 3d space then the classes are ether oddly or evenly divided. If odd, drop the furthest point of the even class from the sole member of the minority class. This problem is now the same as the 3 points in 2d space problem. In the case that there are two points in each class let us assume that there is not any linear function (in 3d space a plane) that can separate the classes. Then by definition all the points must be coplanar. Since we know that the points are not coplanar there must be a plane that can separate the classes.

For the last part I am using the question text from the book which includes “not always” instead of simply “not”, since there are cases where 5 points in 3d space can be separated by a linear function.

If there are n=5 points in 3d space it is possible for them to not be separated by a linear function. Let us say that the positive points are (1,1,1), (4,1,1), and (1,3,1). The negative points are (2,2,0), and (2,2,2). Then the line segment between the negative points directly bisects the region enclosed by the positive ones. Thus any plane that would separate 3d space into a region containing both negative points would also include at least one of the positive ones. Therefore a linear separator cannot always divide two classes of 5 points in 3d space.

5. The initial entrop of Dark Side is:

$I_{\text{dark side}} = -\frac{7}{7+5}\log\frac{7}{7+5} - \frac{5}{7+5}\log\frac{5}{7+5} = .295$

Age

$I_5 = -\frac{2}{2+0}\log\frac{2}{2+0} - \frac{0}{2+0}\log\frac{0}{2+0} = 0$

$I_6 = -\frac{2}{2+1}\log\frac{2}{2+1} - \frac{1}{2+1}\log\frac{1}{2+1} = .276$

$I_7 = -\frac{1}{1+1}\log\frac{1}{1+1} - \frac{1}{1+1}\log\frac{1}{1+1} = .301$

$I_8 = -\frac{1}{1+2}\log\frac{1}{1+2} - \frac{2}{1+2}\log\frac{2}{1+2} = .276$

$I_9 = -\frac{1}{1+1}\log\frac{1}{1+1} - \frac{1}{1+1}\log\frac{1}{1+1} = .301$

E(Age) = 2/12(0)+3/12(.276)+2/12(.301)+3/12(.276)+2/12(.301)=.238

Gain(Age) = .057

Completed Training:

$I_{yes} = -\frac{5}{5+0}\log\frac{5}{5+0} - \frac{0}{5+0}\log\frac{0}{5+0} = 0$

$I_{no} = -\frac{2}{2+5}\log\frac{2}{2+5} - \frac{5}{2+5}\log\frac{5}{2+5} = .260$

E(Completed Training) = 5/12(0)+7/12(.260)=.152

Gain(Completed Training) = .143

Disposition:

$I_{Happy} = -\frac{3}{3+1}\log\frac{3}{3+1} - \frac{1}{3+1}\log\frac{1}{3+1} = .244$

$I_{Sad} = -\frac{2}{2+2}\log\frac{2}{2+2} - \frac{2}{2+2}\log\frac{2}{2+2} = .301$

$I_{Angry} = -\frac{2}{2+2}\log\frac{2}{2+2} - \frac{2}{2+2}\log\frac{2}{2+2} = .301$

E(Disposition) = 4/12(.244)+4/12(.301)+4/12(.301)=.282

Gain(Disposition) = .013

Species:

$I_{Human} = -\frac{3}{3+2}\log\frac{3}{3+2} - \frac{2}{3+2}\log\frac{2}{3+2} = .292$

$I_{Gungan} = -\frac{1}{1+0}\log\frac{1}{1+0} - \frac{0}{1+0}\log\frac{0}{1+0} = 0$

$I_{Wookie} = -\frac{1}{1+1}\log\frac{1}{1+1} - \frac{1}{1+1}\log\frac{1}{1+1} = .301$

$I_{\text{Mon Calamari}} = -\frac{1}{1+1}\log\frac{1}{1+1} - \frac{1}{1+1}\log\frac{1}{1+1} = .301$

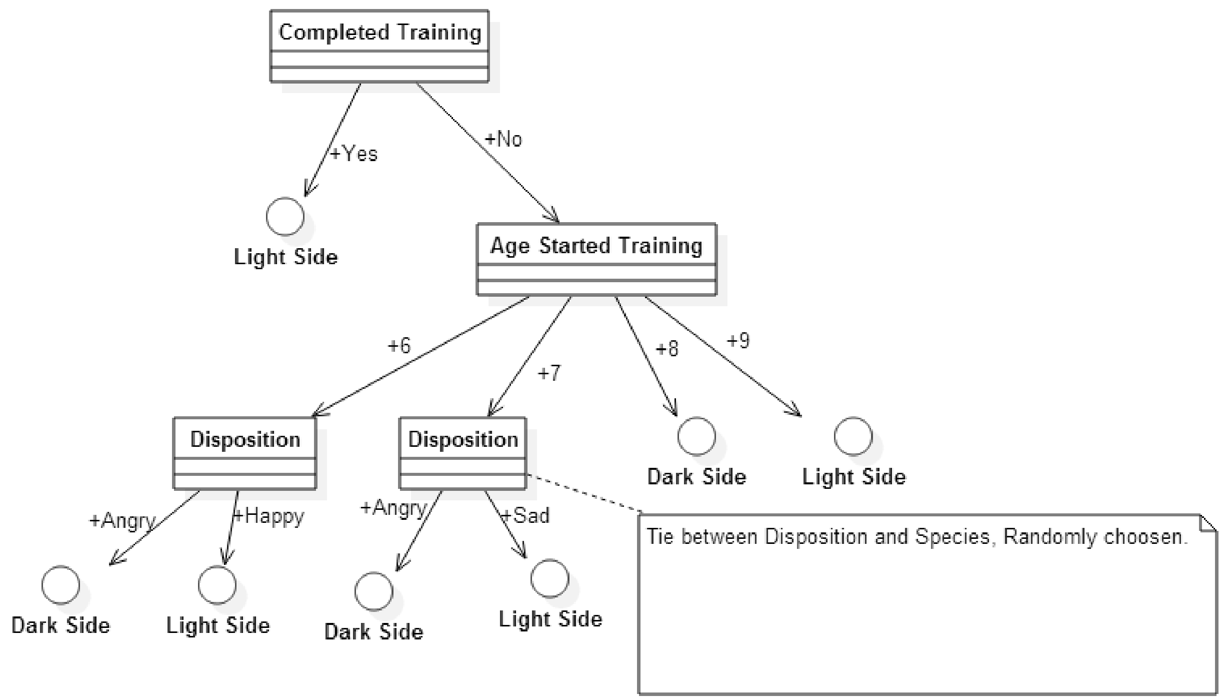
$I_{Ewok} = -\frac{1}{1+1}\log\frac{1}{1+1} - \frac{1}{1+1}\log\frac{1}{1+1} = .301$

E(Species) = 5/12(.292) + 2/12(.301) + 2/12(.301) + 2/12(.301)=.272

Gain(Species) = .023

Logan Esch

Because Completed Training has the greatest information gain(.157), that feature will be the root of the tree.



I would be most confident in predicting that Barbar joins the dark side. This is for many reasons. First, the given data predicts that all who start training at 8 join the dark side unless they complete their training. Second, regardless of age, everyone who did not complete their training and was angry joined the dark side. Even if the training data had some noise, these factors would still give the most confidence as they have the lowest entropy.