**3. Exercise 10**

Exercise 10-1 Information Gain

In this exercise, we want to look more closely at the information gain measure.

Let be a set of training objects with the attributes and the classes .

Let be the disjoint, complete partitioning of produced by a split on attribute (where is the number of disjoint values of ).

(a) Uniform distribution

Compute entropy , entropy for as well as gain given the assumption that the class membership of is uniformly distributed and independent of the values of A. Interpret your result.

**4. Suggested solution:**

Independent uniform distribution

Interpretation: expected - a split on this attribute should not help.

**5. (b) Additional uniform distribution**

We want to analyze how the number of different values influences the information gain. For this, we want to compare two attributes, attribute with values and attribute with values, where the relative frequencies in in values 1 to are identical to that of and in the additional value there is a uniform distribution of the classes. How does differ from Interpret your result.

**6. Suggested solution:**

Interpretation:

In comparison to , for each data object in , we add once and subtract a value , i.e., altogether we add some value

Therefore the information gain must be smaller for compared to .

Thus, would be preferable over for the split, which makes also sense intuitively.

(c) Attributes with many values

Let be an attribute with random values, not correlated to the class of the objects. Furthermore, let have enough values, s.t. not any two instances of the training set share the same value of A. What happens in this situation when building the decision tree? What is problematic with this situation?

**7. Suggested solution:**

Split on : Entropy in each branch is 0 , as we have pure class sets (in each case, some , all others .

Hence we choose as root and the tree is done.

**8. Exercise 10-2 Neurons**

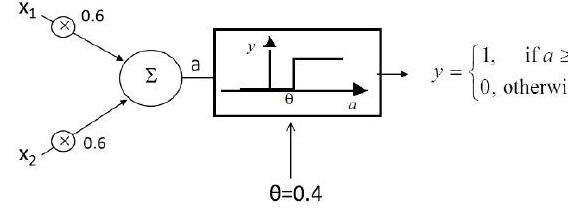
Sketch two trained threshold logic units (that is, individual TLUs, no hidden layer) that can represent for two Boolean variables and the and the OR function, respectively.

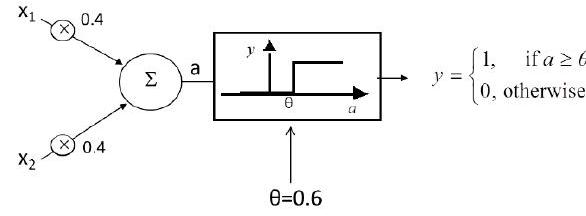
Sketch the related linear separations in the boolean space.

**9. Suggested solution:**

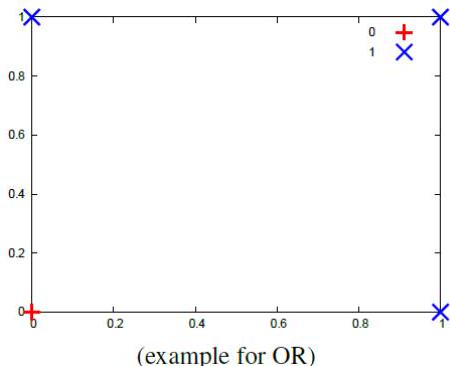
Diverse solutions are possible, examples are:

:





Sketch the related linear separations in the boolean space



for the equations

We have for :

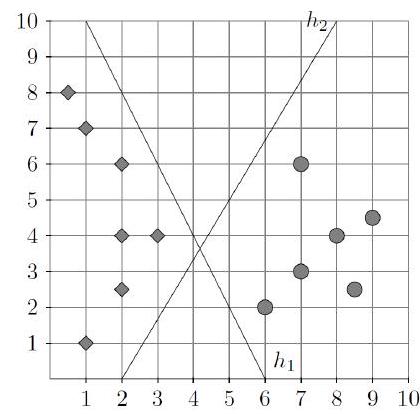
and for :

Or in general:

We have infinitely many other possibilities for separating lines, relating to other weights and thresholds.

Exercise 10-3 Support vectors and margin

Consider the following dataset with points from two classes (diamonds) and (circles).



(a) Give the equations for hyperplanes and .

Suggested solution:

We can start with two points that define a line (i.e., a hyperplane in the two dimensional space). For , we can use (e.g.) and . Thus the slope is

Now using as a point of the line, we get the equation:

For we can take and , thus the slope is

With as a point of the line, we get the equation

(b) Name all the support vectors for and .

**10. Suggested solution:**

The support vectors for are , and .

The support vectors for are and .

(c) Which of the two hyperplanes is better at separating the two classes based on the margin?

Suggested solution:

We compute the margins for the two classifiers by computing the distance from the support vectors to the hyperplanes.

For we have and , such that . Recall that the distance of a point to the hyperplane is .

The distance of support vector is thus:

The distance of support vector is:

The total margin is therefore

For we have and .

The distance of support vector is:

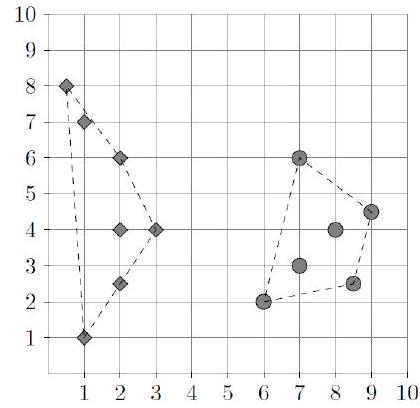
The distance of support vector is:

The total margin is therefore

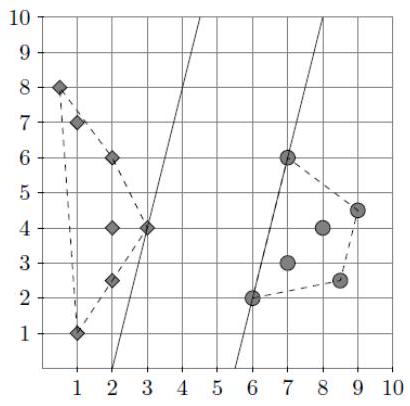
In conclusion, is better than . (d) Find the best separating hyperplane for this dataset, give its equation, and show the corresponding support vectors.

**11. Suggested solution:**

Sketch the convex hull:

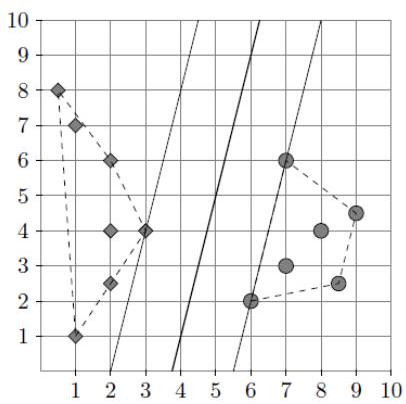


Sketch the maximum possible margin, obvious by the convex hull.



The support vectors are thus and for circles and for diamonds. The optimal hyperplane is therefore

which is exactly half-way between the lines passing through the support vectors (which are and



The margin is .