Gradiance Online Accelerated Learning

07: Machine Learning

Robert

- Home Page
- Handouts
- Tutorials
- Homeworks
- Lab Projects
- Reports
- Class Administration
- Question Bank
- Log Out

Help

Number of questions:	4
Positive points per question:	3.0
Negative points per question:	1.0

Gradiance quiz on Machine Learning. You can attempt to answer the questions as many times as you like. Questions get randomly regenerated each time. The score of the *last* submission gets saved into our records (that is, once you get a perfect score, don't submit again with a bad one).

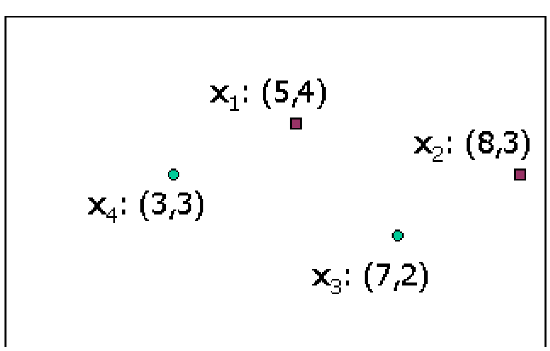
1. We consider the following dataset (+ means positive example and - negative example):

Attributes	+	-
None	32	128
A only	64	256
B only	32	64
A and B	128	64
Total	256	512

For instance, the number of positive examples that have attribute A is 64 and the number of negative example that have neither A or B is 128.
Build a decision tree using information gain as a way of deciding which attribute to use first (there are only 2 attributes here, A and B).
You will need to compute information gain (use log in base 2). Round your answers to the 3rd decimal.
Let us use the following notation: - IG_A (resp. IG_B) = information gain obtained when using A (resp. B) to split the root node - IG_with_A (resp. IG_without_A) = information gain obtained when using B to split data points that have (resp. do not have) attribute A. IG_with_B and IG_without_B are defined exactly in the same way by switching the roles of A and B.

- a) IG_A = 0.012 and IG_B = 0.095
- b) IG_A = 0.879 and IG_B = 0.402.
- c) IG_A = 0.189 and IG_B = 0.255.
- d) IG_with_A = 0.638 and IG_without_A = 0.656.

2. The figure below shows two positive points (purple squares) and two negative points (green circles):



That is, the training data set consists of:

$(\mathbf{x}_1,y_1) = ((5,4),+1)$
$(\mathbf{x}_2,y_2) = ((8,3),+1)$
$(\mathbf{x}_3,y_3) = ((7,2),-1)$
$(\mathbf{x}_4,y_4) = ((3,3),-1)$

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, $\mathbf{x}_2$ and $\mathbf{x}_3$, have a perpendicular bisector that misclassifies $\mathbf{x}_1$ 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.

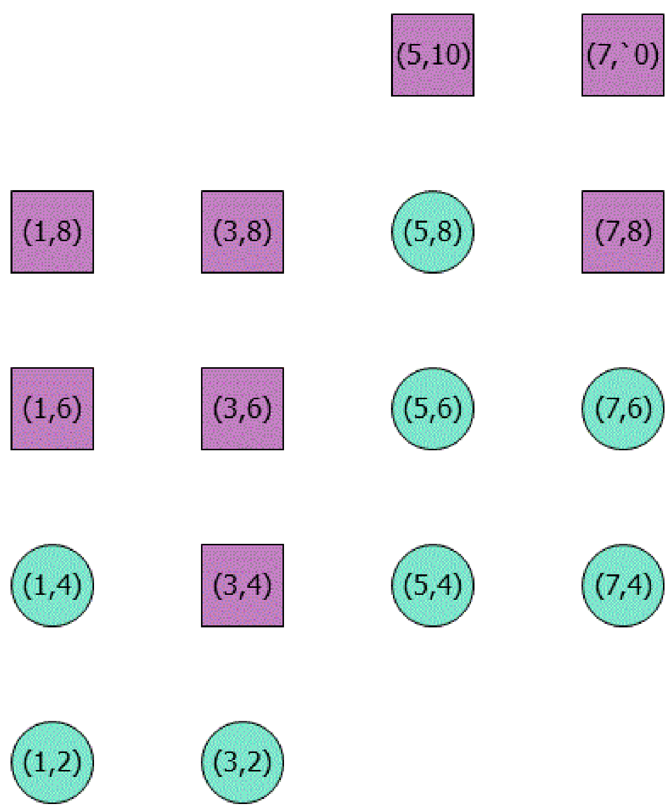
Consider all possibilities for boundaries of this type, and express the boundary as $\mathbf{w}.\mathbf{x}+b=0$, such that $\mathbf{w}.\mathbf{x}+b\geq 1$ for positive points $\mathbf{x}$ and $\mathbf{w}.\mathbf{x}+b\leq -1$ for negative points $\mathbf{x}$. Assuming that $\mathbf{w} = (w_1,w_2)$, identify in the list below the true statement about one of w_1, w_2, and b.

- a) $w_2 = 3/2$
- b) $w_2 = 8/5$
- c) $b = -5$
- d) $w_1 = 1$

3. We consider the following points:
A: (1,5) -
B: (5,8) +
C: (3,4) +
D: (1,3) +
E: (6,-1) -
F: (10,-2) +
G: (8,-4) -
We want to run the perceptron algorithm on them. Specifically, we initialize the weight vector at (0,0) and we consider the stream of points A -> B -> D -> E -> C -> F -> G
We say that the first 4 points are used for training (i.e. to learn w) and the remaining 3 points are used to evaluate performance.
We use the following learning rates: 2, 1, 0.5, 0.25.
Compute the value of w after the training and the misclassification error.

- a) w_final (after 4 steps) = (2,-0.25)
- b) w_final (after 4 steps) = (-2,-1.25) and the classification error is 2/3.
- c) After the 3rd step of training w = (2.5,-0.5) and point E is correctly classified.
- d) After the 2nd step of training w = (3/2, -2).

4. Consider the following training set of 16 points. The eight purple squares are positive examples, and the eight green circles are negative examples.



We propose to use the diagonal line with slope +1 and intercept +2 as a decision boundary, with positive examples above and negative examples below. However, like any linear boundary for this training set, some examples are misclassified. We can measure the goodness of the boundary by computing all the slack variables that exceed 0, and then using them in one of several objective functions. In this problem, we shall only concern ourselves with computing the slack variables, not an objective function.

To be specific, suppose the boundary is written in the form $\mathbf{w}.\mathbf{x}+b=0$, where $\mathbf{w}$ = (-1,1) and b = -2. Note that we can scale the three numbers involved as we wish, and so doing changes the margin around the boundary. However, we want to consider this specific boundary and margin.

Determine the slack for each of the 16 points. Then, identify the correct statement in the list below.

- a) The slack for (7,8) is 0.
- b) The slack for (7,8) is 2.
- c) The slack for (3,6) is 2.
- d) The slack for (5,8) is 0.