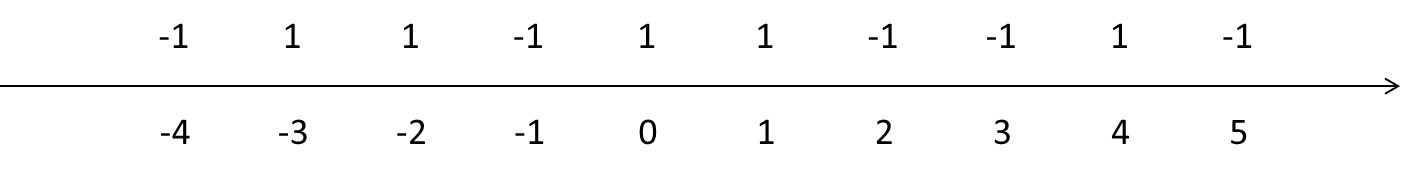
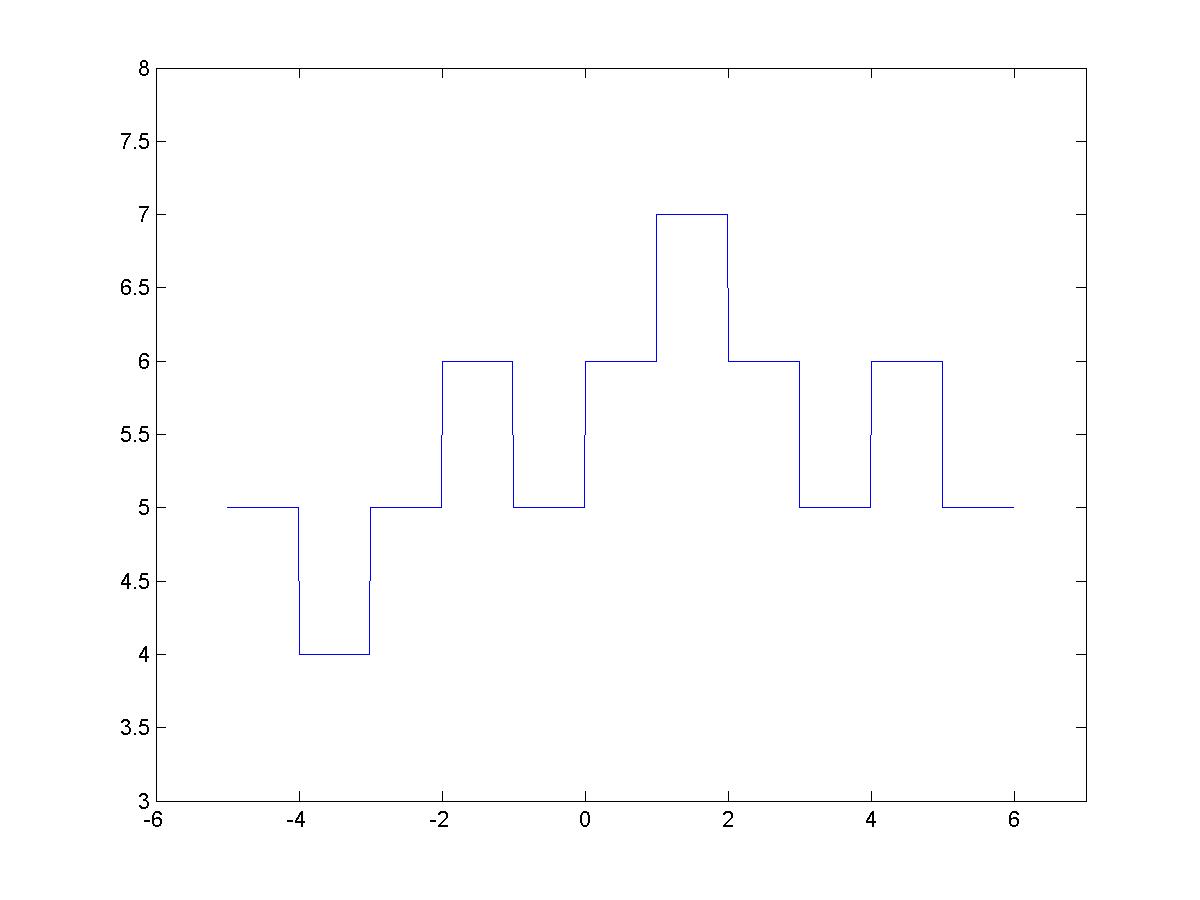
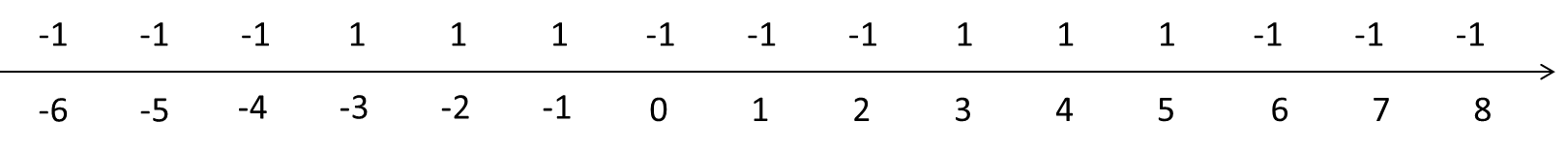
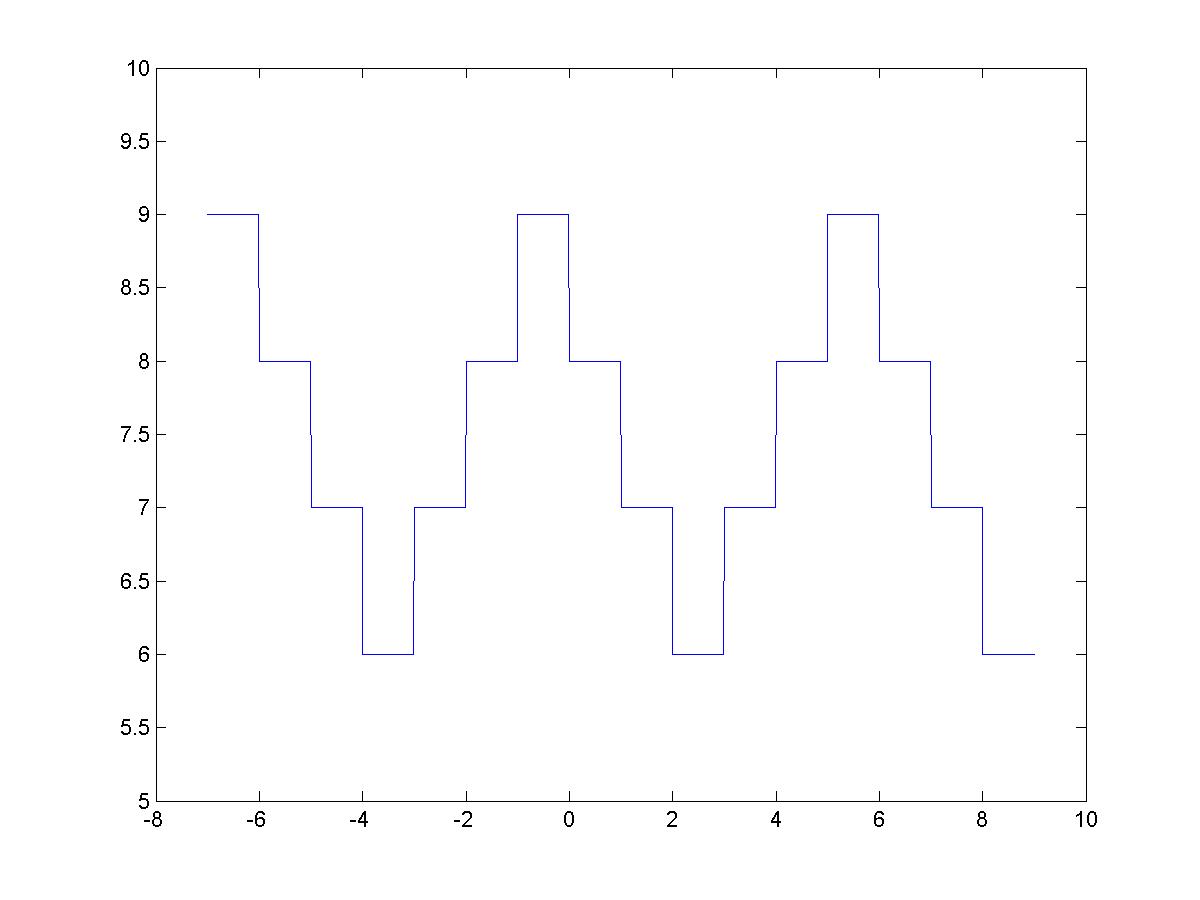
Question I:  
1.  
0-1 loss  
 

2.



* 1. loss



3.

Suppose there are n samples. Pseudo code:

**First calculate the 0-1 loss (record as L) if the decision boundary makes all points +1;**

**L\_tmp = L;**

**Start to move the decision boundary along axis until past the last point:**

**if boundary pasts one point P:**

**if P == -1: L\_tmp = L\_tmp -1;**

**else: L\_tmp = L\_tmp +1;**

**if L\_tmp < L: L = L\_tmp;**

**L is the minimum 0-1 loss;**

This algorithm should only take O(n) since it only traverses the vector once.

4. The hinge loss will give a continuous result instead of discrete result. The problem that’s complicating the algorithm is that the loss value cannot be negative. Therefore for +1 point, f(x)>1 gives 0 loss; for -1 point, f(x)<-1 gives 0 loss. In terms of the algorithm and decision function f(x) = x+b, the algorithm complexity can still be O(n): with the decrease of b, one can know for sure of the x value which makes f(x)>1 (+1 point) or f(x)<-1 (-1 point). Therefore one can just separate two groups, dynamically calculate loss of each group, and compensate the loss by adding some value.

Question 2:  
<http://www.unc.edu/~zyu/ml134/hw1.html>

Hinge loss gives a continuous result, but the result of 0-1 loss is more discrete and radius