Robert Crimi

Machine Learning – Boosting

October 16, 2015

1)

a.

So,

Optimal

b.

So,

Yes, the algorithm assigns the same probability mass function at each round, as the ratio of right to wrong is always 1.

c.

So,

2)

So,

The weak learner assumption states that for a hypothesis h, there is a Dt+1 error that is less than ½. Since our empirical error of the Dt+1 distribution is equal to ½ , we would need to select a different h at round t+1.

3)

Assume

Normalized, assuming m & r are positive:

After another round, both distributions go back to 1/m.

So, at round T, the distributions are both 1/m if T is even and the above distributions if T is odd.