ERM

1) Choose hypothesis space F

2) Learn

If then

For we have three steps:

Step 1 Double Sample Trick

Step 2 Symmetrization

Fix

Where

Step 3 VC-dimension

But we don’t know how it grows actually, maybe exponential or polynomial

when n is small can be exponential but when n is large it is relatively smaller

Thm

Phase change

(To prove it as homework)

So just to focus on what cases cannot be reached

Start with special case: n=d+1, if (0,0…) cannot be realized then

And intuitively it is the worst case so for others it is better

Proof Fix

Note unrealizable patterns as but they maybe intersected. So if I change the first bit from 1 to 0 otherwise keeps the same then the union of them are smaller. So when I replace 1 with 0 then the union is that they cannot have more than d+1 zeros, then it is the special case.

Then d is called the VC-dimension of the set. And d means that there exists can reach every possible results but cannot find

So:

Thm’ with prob over the random draw of training data

Linear classifier: VC-dim = r+1 where

For ERM is the best one and , then with probability

Lecture 4 Practical Algorithms

But indicator function is hard to minimize

So we change classification error from 0-1 loss to other functions.

1. Linear classification

For time limits, if it is a convex optimization and constrain is linear then it is easy to be solved.