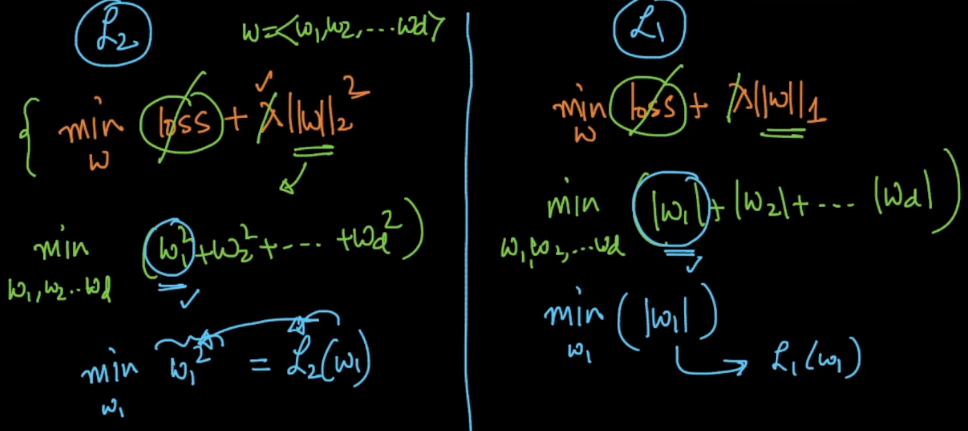


Here we’ll see why L1 regularization create sparsity, means most of w becomes 0.

We’ll compare L2 and L1. Since loss function is same in both reg, so remove them.

Therefore we end up with just w’s, now let’s work with only w1.

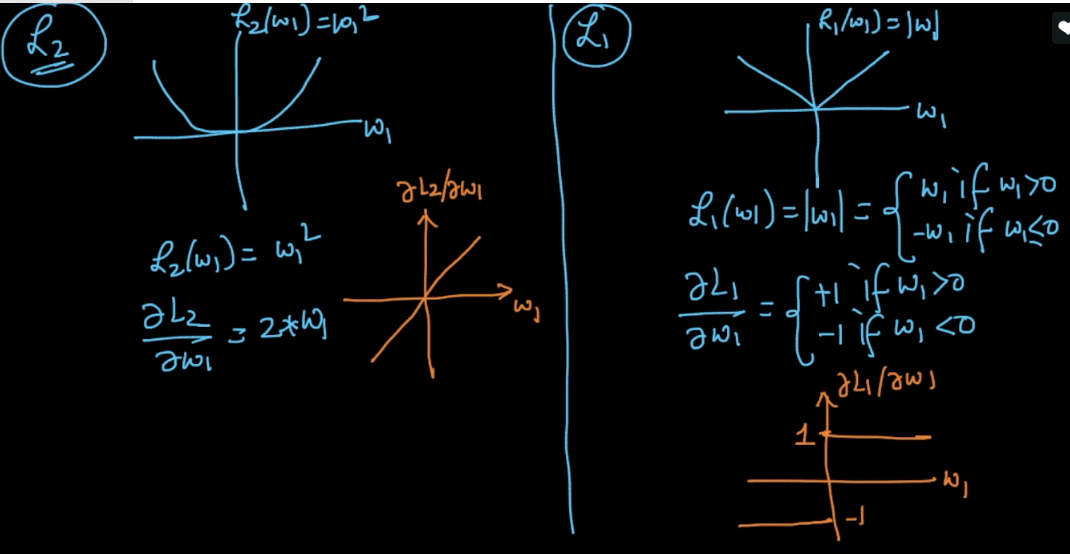
Therefore L2(w1) = min w12 and L1(w1) = min |w1|



Differentiation of L2 = 2 \* w1;

Differentiation of L1 = +1 if w1 > 0 OR -1 if w1 < 0.

Plot graphs for L’s and differentiation of L’s.



Now as we can see that differentiation of L2 is deacreasing with each iteration of w1 value, but of L1 it will be constant either -1 or 1.

Let’s assume w1 at any point j is +Ve.

Therfore for each new w the difference between wj+1 and wj  for L2 is lesser than L1, so L2 will stop before converging to 0, and L1 will converge to 0 for most of w’s.

