Machine Learning - 89511

Assignment 1 - Solution

Boaz Ardel - 203642806

March 2018

Department of Computer Science

Bar-Ilan University

´LaTeX

Contents

Chapter 1

Theoretical Part

1.1 Q1 - General Definition

A Domain - X is a sequence set of bits, for example: $\{0, 1, 1 \ldots, 0\}$, with length of d.
Domain - Y should be single boolean value, therefore can be either 1 or 0 - $\in \{0, 1\}$

B These are all the hypothesis for d=2 - $\{\boldsymbol{x_1, \bar{x_1}, x_2, \bar{x_2}, x_1 \wedge x_2, \bar{x_1} \wedge x_2, x_1 \wedge \bar{x_2}, \bar{x_1} \wedge \bar{x_2}, x_1 \wedge \bar{x_1} \wedge x_2 \wedge \bar{x_2}, \phi}\}$

C There can be 3 states for each var in the hypothesis: appearing, appearing negative and not appearing at all, adding together all these possibilities we would get: $\mathbf{3^d}$, adding the last hypothesis which is our starting point (each var with his negative) - the equation would be: $\mathbf{3^d + 1}$.

D
  i The following example could exist: $\mathbf{\bar{1} \wedge 0 \wedge 1 = 0}$.

  ii This examples can co-exist: $\mathbf{\bar{0} \wedge 1 \wedge 1 = 1}$ (for both).

1.2 Q2 - The Consistency Algorithm

A This algorithm does implement the ERM principle, the ERM principle bring our average error on examples set to minimum. In our case this algorithm's loss function is:
$loss(y, \hat{y}) = \{1, y = 1 \wedge \hat{y} = 0 | 0, y = 0\}$
from this we can infer that in every iteration the updated hypothesis fixed to accept the

current labeled example, thus after halting the average error will be 0. If the average error is 0 then the algorithm is in fact implementing the ERM principle.

B $M(a) \leq 2d$:

Initial hypothesis has $2d$ vars, for each var of the d and his corresponding negative, assuming the algorithm will have a mistake for each iteration then the maximum mistakes will be $2d$ at most.

$M(a) \leq d+1$:

When looking closely at the algorithm we can see that first prediction will always be wrong, then in order to fix the first mistake the algorithm will remove one of the versions of each var (negative or non-negative) leaving the hypothesis with d vars left. Now the maximum mistakes for rest of the iterations will be d, thus the maximum mistakes could be at most $d+1$.

C The run-time would be: $O(d)$, for each example we would process d bits of example set, and checking our hypothesis.

Chapter 2

Practical Part

In attached files as instructed.