Fundamentals of A.I

Surya vardhan Kadiyala,

Id: 11709167

Homework - 5 : supervised machine Learning

Q1) exercise - 7.4 :-

suppose you need to define a system that, given data about a Person TV watching Likes, recommends other TV shows the Person may like. Each show has features specifying whether it is a comedy, whether it features doctors, whether it features Lawyers, & whether it has guns. you are given the fictions examples of fig. 7.23 about whether the Person Like various TV. shows.

example	Comedy	Doctor's	Lawyers	Guns	Likes
e_1	false	true	false	false	false
e_2	true	false	true	false	true
e_3	false	false	true	true	true
e_4	false	false	true	false	false
e_5	false	false	false	true	false
e_6	true	false	false	true	false
e_7	true	false	false	false	false
e_8	false	true	true	true	true
e_9	false	true	true	false	false
e_{10}	true	true	true	false	true
e_{11}	true	true	false	true	false
e_{12}	false	false	false	false	false

b) suppose the error is the sum of absolute errors. Give the optimal decision tree with a squared error?

Sol:- The optimal decision tree, one node predicts Like = false (or no likes)

It has five errors. The sum of squares of errors will be:

$\Rightarrow 5 \times (7/12)^2 + 7 \times (5/12)^2$

$\Rightarrow 1.701 + 1.215$

$\Rightarrow \underline{\underline{2.92}}$

d) suppose the error is the sum of absolute errors. Give the optimal decision tree of depth 2. For each leaf in the tree, give the examples that are filtered with squared error?

Sol:- → The decision tree has depth of '2'.

If lawyer likes = true,

else likes = false.

→ There are '3' errors at the root, & the examples are labeled as {e1, e2, e3 ..... e12}

→ Lawyers who like (true positives) are: {e2, e3, e4, e8, e9, e10} & non-Lawyers who don't like (true-negatives) are: {e4, e5, e6, e7, e11, e12}

→ The probability of a lawyer liking is 3/4, while for non-lawyer it's 5/6.

→ The sum of squares error is calculated as: $2(4/6)^2 + 4(2/6)^2 + (5/6)^2 + 5(1/6)^2$

: $\underline{\underline{2.16}}$

→ In conclusion, the sum of squared error (2.16) is lower than the previous solution (2.92), indicating that this decision tree performs better in classifying the example.

e) what is the smallest tree that correctly calssifies all training exam

Sol:- The smallest decision tree is:

if guns then

(if Lawyers then likes = true,

else likes = false)

else

(if comedy then likes = true

else likes = false)

-> yes, Top-down decision tree will optimize the information gain at each step represent the same function.

f) Give two-instances not appearing in examples of fig-7.23 & show how they are classified using smallest decision tree?

Sol:- smallest - decision - tree:-

if Lawyers = true, then likes = true

else, likes = false.

Two - new examples:

1. comedy = true, doctors = true, Lawyers = false, guns = true
2. comedy = false, doctors = false, lawyers = true, guns = false.

How they are classified:-

1. Likes = false (because not a lawyer)
2. Likes = true (because is a Lawyer).

* Bias explanation:-

The tree only cares if some one is a lawyer. It ignores all other information. This is a biased because:

→ It assumes lawyers always like things.

→ It assumes non-lawyers never like things.

→ It completely ignores other factors like comedy, doctors (or) guns.

Q2) exercise – 7.14

It is possible to define a regularization to minimize $\sum_e (loss(\hat{y}(e)), y(e)) + \lambda * regularizer(\hat{y})$ rather than Formula 7.5 How is this different than the existing regularizer.

Suppose 'λ' is set by k-fold cross validation, & then the model is learned for the whole dataset. How would the algorithm be different for the original ways of defining and regularizers & this alternative way?

Sol:- Part-I:-

→ The Regularizer in Formula 7.5 is designed to minize the sum of errors for each data point. Include a penality term that encourages a simpler model (fewer parameters). this is effective at preventing overfitting on a single dataset.

→ However, if you are working with multiple datasets, the regularizer will still encourage a simpler model. this may not be desirable if you want the model to learn different patterns in each data-set.

→ Alternative approach defines a regularizer that minimizes the sum of errors for each data point. Include a penalty term that encourages a more flexible model.

This regularizer would be more effective at preventing overfitting on multiple datasets using cross-validation.

→ There are a few different ways to define such a regularizer. one option is to use the L1 norm which encourages the model to have few non-zero parameter values. these are just two of many possible options.

→ In, general, the choice of regualizer will depend on the specific problem & data. There is no single best regularizer for all problems. However, for problems where you want the model to be able to learn different patterns in multiple datasets, a regularizer that encourages the model to have more parameters.

Part-2 :-

→ There are few key differences between original regularizer & the alternate regularizer. First, the original regularizer is defined using a single dataset while the alternative regularizer is defined using multiple datasets.

→ The original regularizer encourages the model to be simpler, while the alternative regulizer encourages the model to be more flexible. This may be desirable if we want the model to be able to learn different patterns in each data set.

→ The original regularizer is defing using the L_2 norm while the alternative regularizer is defined using L_1 Norm. This means that the alternative regularizer will encourage the model to have non-zero parameters.

→ The original regularizer is fit on entire dataset, while the alternative regularizer is fit on a subset of the dataset. This means that the alternative regularizer is more effective at preventing overfitting on multiple datasets when using cross-validati-on.

→ There is no single best regularizer for all problems. This choic of regularizer will depend on the specific problem a data. In general, for problems where they want the model to be able to learn different patterns in multiple datasets, a regulizer that encourages the model to have more parameters may be a good choice.

— The End —