Paran Kumar Chebelli
11691712

Fundamentals of AI

Homework 5: Supervised Machine Learning

Question 1) Suppose you need to define a system that, given data about a person's TV watching likes, recommends other TV shows the person may like....

Example	Comedy	Doctors	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_{12}	false	false	false	false	false

Ans)

b) The optimal decision tree, one node predicts like = false (or no likes). It has five errors. The sum of squares of errors will be:

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

$\Rightarrow 1.701 + 1.215$

$\Rightarrow 2.92$

d)

→ the decision-tree has depth of 2

if Lawyer like = true else

likes = false

→ It has 3 errors, At root are all of the example

$(e_1, e_2, e_3, \ldots, e_{12})$

→ Lawyer who like (true positive) are: $\{e_2, e_3, e_4, e_8, e_9, e_{10}\}$

& non-Lawyer who don't like (true negative) are:

$\{e_1, e_5, e_6, e_7, e_{11}, e_{12}\}$

→ the probability of a Lawyer liking is 3/4, while for non-Lawyer's its 5/6

→ the sum of squares error is:

$$= 2\left(\frac{4}{6}\right)^2 + 4\left(2/6\right)^3 + \left(5/6\right)^2 + 5\left(1/2\right)^2$$

$$= 2.16$$

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

e) What is the smallest tree that correctly classifies all training example?

Ans)

The smallest decision-tree is

if genres then

(if Lawyer 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 example of fig 7.23 & show how they are classified using smallest decision tree?

dd) Smallest decision tree:

if lawyer = true, then likes = true
else, likes = false.

Two new examples:

1- Comedy = true, doctor = true, Lawyer = false, guns = true
2- Comedy = false, doctor = false, lawyer = 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 someone is a lawyer. It ignores all other information. This is 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 7.10.14

It is possible to define a regularizer to minimize $\sum_e (error_h(e) + \lambda^* regularizer_h)$ rather than Formula 7-5 ....

Ans) Part 1

→ The Regularizer in formula 7-5 is designed to minimize the sum of errors for each data point plus a penalty term that encourages the model to be simpler (i.e. to have fewer parameters). This regularize is effective at preventing overfitting on a single dataset.

→ However, if you are working with multiple datasets, the regularizer will encourage the model to be simpler (i.e. to have fewer parameters). This may not be desirable if you want the model to be able be learn different patterns in each dataset.

→ Alternatively, you could define a regularizer that minimize the sum of the error for each data point plus a penalty term that encourages the model to be more flexible (i.e. to have more parameters). The regularizer would be more effective at preventing overfitting on multiple datasets (a) when using cross validation.

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

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

Part 2:

→ there are few key difference between the original regularizer & the alternative 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 regularizer 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 dataset.

→ The original regulaizer is defining using the L_2 Norm while the alternative regulaizer is defined using L_1 Norm. this means that the alternative regulaizer will encourage the model to have non-zero parameter.

→ the original regulaizer is fit on the entire dataset, while the alternative regulaizer is fit on a subset of the dataset. this means that the alternatives regulaizer is more effective at presenting overfitting on multiple datasets when using cross validation.

→ there is no single best regulaizer for all problems. the choice of regulaizer will depends on the specific problems data. In general, for probleys where they want the model to be able to learn different patterns in multiple datasets, a regulaizer that encourages the model to have more parameters may be a good choice.