Fundamentals of Artificial Intelligence

HomeWork - 5
CSCE - 5210

Ajay Reddy Yedumula
11580520

Q1. Suppose, We have a System that observes a person's TV Watching habits in Order to recommend other TV Shows the person may like Suppose that We have a characterized each show by Whether it's a Comedy, Whether it's features doctors. Whether it features lawyers and Whether it has guns, Suppose We are given the examples of fig 7.23 about Whether person likes Various TV Shows.

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_6	true	false	false	true	false
e_7	true	false	false	false	true
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

We Want to use this dataset to learn the Value of Likes (i.e., to predict which TV shows the person would like based on the attribute of TV Show)

b) Do the same as in part (a) but with Sum of Squares errors

Ans:- The optimal decision tree with one node products likes = false (or not like). It has 5 errors then the Sum of Square of errors will be

$$5 \times \left(\frac{7}{12}\right)^2 + 7 * \left(\frac{5}{12}\right)^2 = 1.70 + 1.215 = 2.92.$$

d) Do the Same as in part (c), but with the Sum of Square of errors.

Ans:- An optimal solution with depth 2 is if lawyers then likes = true else like = false.
It has 3 errors, At root are all of examples $(e_1 \ldots e_{12})$ filtered to lawyer true node are $\{e_2, e_3, e_4, e_8, e_9, e_{10}\}$ filtered to lawyer false node are $\{e_1, e_5, e_6, e_7, e_{11}, e_{12}\}$ if lawyer then likes= 4/6 els like = 3/6.
The Sum of Square errors is

$$2\left(\frac{4}{6}\right)^2 + 4\left(\frac{2}{6}\right)^3 + 7\left(\frac{5}{6}\right)^2 + 5\left(\frac{1}{6}\right)^2 = 2.16$$

e) What is the Smallest tree that correctly classifies all training examples? Does a top-down decision tree that optimize the information gain al each step represents the Same function.

Ans The Smallest decision tree is if guns then (if lawyers then likes = true else like = false)
else
(if Comedy then likes = true else likes = false)

One way to find such trees is to do a two-step look ahead for each property, check a Split on that property and then do a split on each leaf before evaluating the split.

f) Give two instance not appearing in the examples of fig F23 and show how they are classified using the Smallest decision trees use this to explain the bias inherit in the tree (How does the bias give you these particular predictions)

Ans: The logistic regression learns algorithm can learn any linearly Separately classification. The error can be made arbitarly. Small for arbitary sets of examples if and only if the target classification is linearly Separable.

Q.2) It is possible to define a regularizer to minimize Σ_e (error$_h$(e) + λ * regularize$_h$) rather than formula 7.5 How this is different than the existing regularizer? [Hint: Think about this effects multiple datasets or for cross validation].

Suppose λ is a set by k-fold cross validation, and then the model is learned for the whole dataset. How would the algorithm be different from the original ways of defining a regularizer and this alternative way? [Hint: There is a different number of examples used for regularization than there is the full dataset; does this matter?] Which works better in practice?

Ans (I) The regularizer in formula 7.5 is designed to minimize the sum of the error for each data point plus a penalty term that encourages the model to be simpler (i.e., to have fewer parameters). This regularizer is effective at presenting overfitting on a single data set.

(II) The alternative regularizer is more effective at presenting overfitting on presenting multiple dataset or when using cross-validation for problems. Where we want the model to have more parmaters may be a good choice.

Explanation:

(I) If you are working with multiple datasets or using cross-validation. This regularizer may not be ideal. In particular, If the model is fit on multiple datasets, the regularizer will encourage the model to be simpler. This may not be desirable

if you want the model to be able to learn different patterns in each dataset.

⇒ Alternatively, you could define a regularizer that minimizes the sum of the errors. For each data point plus a penalty term that encourages the model to be more flexible. This regularities would be more effective at preventing over fitting on multiple datasets or when using Cross-Validation

⇒ There are a few different Way to define such a regularities one option is to use the a_1 norm which encourages the model to have small parmeter values. These are just two of many possible options.

⇒ In general, the choice of regularities will depend on the specific problem and data. There is no single best regularized for all problems. However, for more problems Where you Want the model to be able to learn different patterns in multiple datasets. A regularities that encourages the model to have more parameters may be a good choice.

(II) ⇒ There are few key differences between the orginal regularizes and the Alternative regularities. First, the original regularizer is defined using a single dataset, While the Alternative regularizer is defined using multiple datasets. This means that the Alternative regularizer will be more effective at preventing over fitting on multiple datasets.

⇒ Second, the original regularizer encourages the model to be simpler (i.e., to have fewer parameters). 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.

⇒ Third, the original regularizer is defined using the l_2 norm, while the alternative regularizer's defined using L_1 norm. this means that the alternative regularizers will encourage the model to have fewer non-zero parameters.

⇒ Fourth, the original regularizer is fit to the entire dataset, while, the alternative regularizer is fit on the subset of dataset. This means that the alternative.

⇒ Regularities will be less likely to overfit on the data. Overall, the alternative regularizer is more effective at preventing overfitting on multiple datasets or when using cross-validation.

⇒ There is no single best regularities for all problems the choice of regularities will depend on the specific problem and data in general. For problems where they want the model to be able to learn different patterns in multiple datasets. A regularizer that encourages the model to have more parameters may be a good choice.