

Homework Assignment #4

Assigned: Sunday 04/19/2015; Due: Tuesday 04/28/2015 11:59pm (via Onourse).

(total: 80 points)

Problem 1. (10 points) Give decision trees to represent the following Boolean functions

- a) (2 points) $A \wedge \bar{B}$
- b) (2 points) $A \vee [B \wedge \bar{C}]$
- c) (2 points) $\bar{A} \oplus B$
- d) (2 points) $[A \wedge B] \vee [C \wedge D]$
- e) (2 points) $[A \vee B] \wedge [C \vee \bar{D}] \oplus B$

where \bar{A} is a negation of A and \oplus is an exclusive OR operation.

Problem 2. (10 points) Consider the data set from the following table

	<i>Sky</i>	<i>Temperature</i>	<i>Humidity</i>	<i>Wind</i>	<i>Water</i>	<i>Forecast</i>	<i>Enjoy Sport</i>
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

- a) (5 points) Using *Enjoy Sport* as the target, show the decision tree that would be learned if the splitting criterion was information gain.
- b) (5 points) Add the training example from the table below and compute the new decision tree. This time, show the value of the information gain for each candidate attribute at each step in growing the tree.

	<i>Sky</i>	<i>Temperature</i>	<i>Humidity</i>	<i>Wind</i>	<i>Water</i>	<i>Forecast</i>	<i>Enjoy Sport</i>
1	Sunny	Warm	Normal	Weak	Warm	Same	No

Problem 3. (20 points) Consider a two-layer feed-forward neural network with two inputs x_1 and x_2 , one hidden unit z , and one output unit f . This network has five weights ($w_{1z}, w_{2z}, w_{0z}, w_{zf}, w_{0f}$), where w_{0z} represents the bias for unit z and where w_{0f} represents the bias for unit f . Initialize these weights to the values $(0.1, -0.1, 0.1, -0.1, 0.1)$, then give their values after each of the first two training iterations of the backpropagation algorithm. Assume learning rate $\eta = 0.3$, momentum $\mu = 0.9$, incremental weight updates, and the following training examples

$$\{(x_1, x_2)_i, y_i\}_{i=1}^2 = \{(1, 0), 1\}, \{(0, 1), 0\}.$$

Problem 4. (10 points) Consider minimizing the following regularized error function using the backpropagation algorithm

$$E = \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - f_{ij})^2 + \gamma \sum_{i,j} w_{ij}^2$$

Derive the gradient descent update rule for this error function. Show that it can be implemented by multiplying each weight by some constant before performing the standard gradient descent update as shown in class.

Problem 5. (30 points) Using the neural network code from class as a starting point, compare the following ensemble methods in a binary classification scenario: 30 bagged neural networks with 30 bagged regression trees. In each case, the final decision is made by averaging the raw outputs from the ensemble members. Use at least 3 different classification data sets to provide comparisons. Select appropriate data sets from the UCI Machine Learning Repository. Proper comparisons should be made using 10-fold cross-validation experiments. Use your knowledge about model comparison to formally conclude which of the two algorithms is better. Normalize the variables on the training set, then normalize the test set using the normalization parameters from the training set (e.g. mean and standard deviation for each feature). If data sets have categorical variables, provide a proper approach to convert them into numerical variables. Use classification error to evaluate quality of classifiers.

Homework policies (read carefully):

Your assignment must be typed; for example, in Latex, Microsoft Word, Lyx, etc. Images may be scanned and inserted into the document if it is too complicated to draw them properly. Submit a single pdf document or if you are attaching your code submit your code together with the typed (single) document as one .zip file.

All code (if applicable) should be turned in when you submit your assignment. Use Matlab, Python or R.

Policy for late submission assignments: Unless there are legitimate circumstances, late assignments will be accepted up to 5 days after the due date and graded using the following rule:

- on time: your score \times 1
- 1 day late: your score \times 0.9
- 2 days late: your score \times 0.7

3 days late: $\text{your score} \times 0.5$

4 days late: $\text{your score} \times 0.3$

5 days late: $\text{your score} \times 0.1$

For example, this means that if you submit 3 days late and get 80 points for your answers, your total number of points will be $80 \times 0.5 = 40$ points.

All assignments are individual, except when collaboration is explicitly allowed. All the sources used for problem solution must be acknowledged, e.g. web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously; for detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.

Good luck!