

# CSCI 5512: Artificial Intelligence II

Spring 2011

## Homework #3, Due May 2nd

1. (20 points) This question considers the value of perfect information (VPI)

$VPI(E_j)$  which evaluates the value of additional information  $E_j$  given existing information  $E$ .

(a) (8 points) Show that VPI is non-negative, i.e.,  $VPI(E_j) \geq 0, \forall j, E$ .

(b) (12 points) Show that VPI is order independent. Given two pieces of information, the value of both is independent of the order that it comes in.

2. (20 points) The following payoff matrix shows a game between politicians and the Federal Reserve. Politicians can expand or contract fiscal policy, while the Fed can expand or contract monetary policy. Each side also has preferences for who should do what. The payoffs shown are simply the rank orderings: 9 for first choice through 1 for last choice.

	Fed: Contract	Fed: Do Nothing	Fed: Expand
Pol: Contract	F = 7, P = 1	F = 9, P = 4	F = 6, P = 6
Pol: Do Nothing	F = 8, P = 2	F = 5, P = 5	F = 4, P = 9
Pol: Expand	F = 3, P = 3	F = 2, P = 7	F = 1, P = 8

(i) (12 points) Find the Nash equilibrium for the game in pure strategies.

(ii) (4 points) Is there a dominant strategy equilibrium in the game? If yes, find the dominant strategy equilibrium; if no, explain why the Nash equilibrium is not also a dominant strategy equilibrium.

(iii) (4 points) We say that an outcome is Pareto optimal if there is no other outcome that all players would prefer. Is the pure strategy Nash equilibrium Pareto optimal?

3. (40 points) (Programming Assignment) Consider the Restaurant dataset given in the table below:

Example	Attributes										Target
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	WillWait
X <sub>1</sub>	T	F	F	T	Some	\$\$\$	F	T	French	0–10	T
X <sub>2</sub>	T	F	F	T	Full	\$	F	F	Thai	30–60	F
X <sub>3</sub>	F	T	F	F	Some	\$	F	F	Burger	0–10	T
X <sub>4</sub>	T	F	T	T	Full	\$	F	F	Thai	10–30	T
X <sub>5</sub>	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X <sub>6</sub>	F	T	F	T	Some	\$\$	T	T	Italian	0–10	T
X <sub>7</sub>	F	T	F	F	None	\$	T	F	Burger	0–10	F
X <sub>8</sub>	F	F	F	T	Some	\$\$	T	T	Thai	0–10	T
X <sub>9</sub>	F	T	T	F	Full	\$	T	F	Burger	>60	F
X <sub>10</sub>	T	T	T	T	Full	\$\$\$	F	T	Italian	10–30	F
X <sub>11</sub>	F	F	F	F	None	\$	F	F	Thai	0–10	F
X <sub>12</sub>	T	T	T	T	Full	\$	F	F	Burger	30–60	T

(a) (15 points) Implement a decision tree learning algorithm `dtree4`, and learn a decision tree (of depth at most 4) from the given training data. What is the training set error of the learned decision tree?

(b) (15 points) Implement a 2-decision list learning algorithm `dlist2`, and learn a decision list from the given training data. What is the training set error of the learned decision tree?

(c) (10 points) Show that the hypothesis space of  $k$ -decision lists include all decision trees of depth  $k$ .

4. (40 points) (Programming Assignment) We will use the Pendigits dataset from the UCI machine learning repository. The dataset contains 10992 examples, each with 16 integer features in the range 0-100 and a unique class label in  $\{0,1,2,\dots,9\}$ . The problem is a 10-class classification problem where a class corresponds to a handwritten number.

The training set contains 7494 data points and can be downloaded in plain text from:

<http://archive.ics.uci.edu/ml/machine-learning-databases/pendigits/pendigits.train>

The test set contains 3498 data points and can be downloaded in plain text from:

<http://archive.ics.uci.edu/ml/machine-learning-databases/pendigits/pendigits.test>

Each data point in the training as well as test set files is a line of 17 comma-separated-values—the first 16 contains the input feature vector and the last one is the class label. For the training set, the class label is to be used for training whereas for the test set, the class label is to be only used for evaluation. You can find more information about the dataset from:

<http://archive.ics.uci.edu/ml/datasets/Pen-Based+Recognition+of+Handwritten+Digits>

Train a feed-forward neural network nnet with one hidden layer using the backpropagation algorithm. Choose the number of hidden nodes appropriately. Set the number of output nodes to 10, one for each class. Using 10 different initial choices for the starting weights, report:

- (i) Mean accuracy on the test set along with standard deviation,
- (ii) Accuracy on the test set using the initial choice of weights which performed best on the training set
- (iii) Best accuracy on the test set among all 10 initial choices of weights.