IS5152 Data-driven Decision Making
SEMESTER 2 2023-2024
Assignment 4

- Upload a softcopy of your solution as a **pdf** document to Canvas.
- Deadline: 11.59 pm, Friday, 12 April, 2024.

---

1. (10 points) We have built a classifier using a data set consisting of 15 data samples. The following table shows the target variable y_i, the three input variables x_1, x_2, x_3 and the predicted output p_i from the classifier:

i	y_i	x_1	x_2	x_3	p_i
1	1 = YES	0	1	35	0.8227
2	0 = NO	0	1	21	0.0034
3	1 = YES	0	0	54	0.7462
4	1 = YES	0	0	62	0.9945
5	0 = NO	1	1	88	0.7363
6	0 = NO	1	1	26	0.0000
7	0 = NO	1	1	73	0.0012
8	1 = YES	0	1	47	0.9996
9	1 = YES	1	1	84	0.2625
10	0 = NO	0	0	49	0.1829
11	0 = NO	1	1	20	0.0000
12	0 = NO	0	0	48	0.1180
13	0 = NO	0	1	29	0.1743
14	1 = YES	0	0	58	0.9584
15	0 = NO	0	0	19	0.0000

(a) (2 points) We make prediction according to this rule: "If prediction $p_i \geq$ threshold, then Class 1, else Class 0." What threshold value would you choose to maximize the accuracy of this rule on the 15 training data samples?

Sort the data according to the predicted output values:

```
 #6  0.0000   NO
#11  0.0000   NO
#15  0.0000   NO
 #7  0.0012   NO
 #2  0.0034   NO
#12  0.1180   NO
#13  0.1743   NO
#10  0.1829   NO
 #9  0.2625   YES
 #5  0.7363   NO
 #3  0.7462   YES
 #1  0.8227   YES
```

```
#14  0.9584   YES
 #4  0.9945   YES
 #8  0.9996   YES
```

Let threshold be

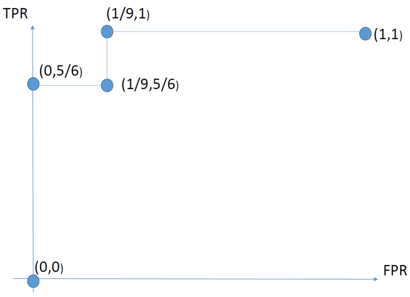
- $(0.1829+0.2625)/2 = 0.2227$ or
- $(0.7363+0.7462)/2 = 0.74125$

to obtain highest accuracy of $14/15 = 0.9333$

(b) (4 points) Compute the precision and recall, as well as the F-measure value (with $\beta = 1$) of the model using the threshold value you determine in part (a) above.

- Threshold = 0.2227
  - precision = 6/7, recall = tpr = 6/6 = 1,
  - F-measure = $(2 \times tpr \times prc)/(prc + tpr) = 2 \times 1 \times (6/7)/(13/7)$ $= 12/13 = 0.9231$
- Threshold = 0.74125
  - precision = 5/5 = 1, recall = tpr = 5/6
  - F-measure = $(2 \times tpr \times prc)/(prc + tpr) = 2 \times (5/6) \times 1/(11/6)$ $= 10/11 = 0.9091$

(c) (4 points) Compute the area under the ROC curve of this classifier.



- Threshold = 0.2227, TPR = 1, FPR = 1/9.
- Threshold = $(0.2625+0.7363)/2$, TPR = 5/6, FPR = 1/9.
- Threshold = $(0.7363+0.7462)/2$, TPR = 5/6, FPR = 0.
- Total area = $(5/6 \times 1/9) + (1 \times 8/9) = 53/54 = 0.9815$

2. (10 points) A system analyst studied the effect of computer programming experience on ability to complete within specified time a complex programming task. Ten persons who had varying amount of experience (in months) were selected for the study. All persons were given the same programming task, and the results of their success in the task are shown in the table below. The results are coded in binary fashion: $d = 1$ if the task was completed successfully in the alloted time, $d = 0$ otherwise. We are interested in building a model to predict if the given task can be completed successfully within the alloted time using the amount of experience (in months) as the only input.

Person	1	2	3	4	5	6	7	8	9	10
Experience (months)	14	15	6	8	29	10	25	12	30	18
Task success	0	0	0	0	1	1	1	1	1	1

For the questions below, consider only binary decision trees, that is, any node in the trees can have at most two branches.

(a) (2 points) What is the entropy of the training samples with respect to the classification?

First sort the data according to the attribute "Experience":

$Entropy(S) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 = -0.4 \log_2 0.4 - 0.6 \log_2 0.6 = 0.97095.$

(b) (3 points) Build a complete decision tree by maximizing the information gain at each node split.

Sort the data according to Experience. There are 3 possible splits:

Person	3	4	6	8	1	2	10	7	5	9
Experience (months)	6	8	10	12	14	15	18	25	29	30
Task success	0	0	1	1	0	0	1	1	1	1

- Experience ≤ 9 versus Experience > 9:
  Information gain

$$
\begin{aligned}
&= 0.97095 - \frac{2}{10}(0) - (\frac{8}{10})(-\frac{2}{8}\log_2\frac{2}{8} - \frac{6}{8}\log_2\frac{6}{8}) \\
&= 0.97095 - (\frac{8}{10})(0.811278) = 0.97950 - 0.64902 = 0.32193
\end{aligned}
$$

- Experience ≤ 13 versus Experience > 13:
  Information gain

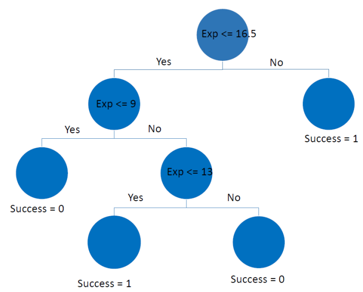
$$
\begin{aligned}
&= 0.97095 - \frac{4}{10}(1) - (\frac{6}{10})(-\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6}) \\
&= 0.97095 - \frac{4}{10} - \frac{6}{10}(0.918295) = 0.97950 - 0.95098 = 0.019973
\end{aligned}
$$

- Experience ≤ 16.5 versus Experience > 16.5:
  Information gain

$$
\begin{aligned}
&= 0.97095 - (\frac{6}{10})(-\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6}) - (\frac{4}{10})0 \\
&= 0.97095 - (\frac{6}{10})(0.918295) = 0.97950 - 0.55078 = 0.41998
\end{aligned}
$$

The third split at 16.5 is the best. The decision tree/rule is:

```
If Experience <= 16.5, then
     if Experience <= 9, then Task success = 0
     else if Experience <= 13, then Task success = 1
                                else Task success = 0
Else
     Task success = 1
```



(c) (2 points) Compute the GINI diversity index for the given data samples.
Gini index = $2p_1p_2 = 2 \times 0.4 \times 0.6 = 0.48$.

(d) (3 points) Build a second complete decision tree where at each node split, the reduction in node impurity as measured by the GINI diversity index is maximized. There are 3 possible splits:

- Experience ≤ 9 versus Experience > 9:
  Gini index = $0.2 \times 0 + 0.8 \times (2 \times \frac{2}{8} \times \frac{6}{8}) = 0.3$
- Experience ≤ 13 versus Experience > 13:
  Gini index = $0.4 \times (2 \times 0.5 \times 0.5) + 0.6 \times (2 \times \frac{2}{6} \times \frac{4}{6}) = 0.2 + \frac{4}{15} = 0.4667$
- Experience ≤ 16.5 versus Experience > 16.5:
  Gini index = $0.6 \times (2 \times \frac{2}{6} \times \frac{4}{6}) + 0.4 \times 0 = \frac{4}{15} = 0.2667$

The third split results in the lowest Gini diversity index. The decision tree is the same as in part (b)

Sample no.	A1 (in $000)	A2 (in $000)	A3 (= constant)	Loan approved?
1	8	2.0	1	no
2	1	0.5	1	no
3	4	6.0	1	yes
4	2	2.5	1	yes

3. (10 points) Suppose you have a loan application data set that consists of just two attributes: A1. amount of credit requested and A2. Average monthly saving. The samples in the data set are as follows:

   A target output $t_i = 1$ is assigned for a loan that is approved, and $t_i = -1$ is assigned for a loan that is not approved, $i = 1, 2, 3, 4$. A single neuron is to be trained to distinguish the two groups of samples.

   (a) (5 points) The neuron is trained using the perceptron learning rule starting from an initial weight

   $$\mathbf{w^0} = \begin{pmatrix} -1 \\ 5 \\ 1 \end{pmatrix}$$

   Is there any value for the learning constant η that would give us a solution of the problem immediately after the first sample is presented?

   $$\begin{aligned} \mathbf{x^1 w^0} &= \begin{pmatrix} 8 \\ 2 \\ 1 \end{pmatrix}^t \begin{pmatrix} -1 \\ 5 \\ 1 \end{pmatrix} = 3 \\ \mathbf{w^1} &= \mathbf{w^0} - 2\eta \mathbf{x^1} \\ &= \begin{pmatrix} -1 \\ 5 \\ 1 \end{pmatrix} - 2\eta \begin{pmatrix} 8 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1-16\eta \\ 5-4\eta \\ 1-2\eta \end{pmatrix} \end{aligned}$$

   Find η such that

   $$\begin{aligned} (\mathbf{x^1})^\mathbf{t}\mathbf{w^1} &= \begin{pmatrix} 8 \\ 2 \\ 1 \end{pmatrix}^t \begin{pmatrix} -1-16\eta \\ 5-4\eta \\ 1-2\eta \end{pmatrix} = 3 - 138\eta < 0 \\ (\mathbf{x^2})^\mathbf{t}\mathbf{w^1} &= \begin{pmatrix} 1 \\ 0.5 \\ 1 \end{pmatrix}^t \begin{pmatrix} -1-16\eta \\ 5-4\eta \\ 1-2\eta \end{pmatrix} = 2.5 - 20\eta < 0 \\ (\mathbf{x^3})^\mathbf{t}\mathbf{w^1} &= \begin{pmatrix} 4 \\ 6 \\ 1 \end{pmatrix}^t \begin{pmatrix} -1-16\eta \\ 5-4\eta \\ 1-2\eta \end{pmatrix} = 27 - 90\eta > 0 \\ (\mathbf{x^4})^\mathbf{t}\mathbf{w^1} &= \begin{pmatrix} 2 \\ 2.5 \\ 1 \end{pmatrix}^t \begin{pmatrix} -1-16\eta \\ 5-4\eta \\ 1-2\eta \end{pmatrix} = 11.5 - 44\eta > 0 \end{aligned}$$

   Answer: $0.125 < \eta < \frac{23}{88}$

   (b) (5 points) Suppose the neuron is now trained using gradient descent

- starting from the same initial weight
- to minimize the sum of the errors:

$$\frac{1}{2}\sum_{i=1}^{4}(t_i - o_i)^2$$

where the linear output o_i is computed as:

$$o_i = \mathbf{w^t x_i}$$

- let the learning rate $\eta = 0.05$.

Find $\mathbf{w^1}$.

i	t_i	o_i	$t_i - o_i$
1	-1	(8)(-1) + (2)(5) + (1)(1) = 3	-4
2	-1	(1)(-1) + (0.5)()5) + (1)(1) = 2.5	-3.5
3	1	(4)(-1) + (6)(5) + (1)(1) = 27	-26
4	1	(2)(-1) + (2.5)(5) + (1)(1) = 11.5	-10.5

$$\begin{aligned}
\sum_{i=1}^{4}(t_i - o_i)\mathbf{x}_i &= -4\begin{pmatrix} 8 \\ 2 \\ 1 \end{pmatrix} - 3.5\begin{pmatrix} 1 \\ 0.5 \\ 1 \end{pmatrix} - 26\begin{pmatrix} 4 \\ 6 \\ 1 \end{pmatrix} - 10.5\begin{pmatrix} 2 \\ 2.5 \\ 1 \end{pmatrix} = \begin{pmatrix} -160.5 \\ -192 \\ -44 \end{pmatrix} \\
\mathbf{w^1} &= \mathbf{w^0} - \eta\nabla\mathbf{E} \\
&= \mathbf{w^0} - \eta(\frac{1}{2})\sum_{i=1}^{4}(2)(t_i - o_i)(-\mathbf{x}_i) \\
&= \mathbf{w^0} + \eta\sum_{i=1}^{4}(t_i - o_i)\mathbf{x}_i \\
&= \begin{pmatrix} -1 \\ 5 \\ 1 \end{pmatrix} + 0.05\begin{pmatrix} -160.5 \\ -192 \\ -44 \end{pmatrix} = \begin{pmatrix} -9.025 \\ -4.6 \\ -1.2 \end{pmatrix}
\end{aligned}$$