

Due date and time: Dec. 9th 2:30pm

Submit in class (hardcopy), use blackboard, or e-mail me (sael@sunykorea.ac.kr).

* **points max**

1. pg. 763 Excercise18.8 [10 points] Decision Trees

Consider the following data set comprised of three binary input attributes (A_1, A_2 , and A_3) and one binary output:

Example	A_1	A_2	A_3	Output y
x_1	1	0	0	0
x_2	1	0	1	0
x_3	0	1	0	0
x_4	1	1	1	1
x_5	1	1	0	1

Use the DECISION-TREE-LEARNING algorithm to learn a decision tree for these data. Show the computations made to determine the attribute to split at each node.

```

function DECISION-TREE-LEARNING(examples, attributes, parent_examples) returns
a tree

if examples is empty then return PLURALITY-VALUE(parent_examples)
else if all examples have the same classification then return the classification
else if attributes is empty then return PLURALITY-VALUE(examples)
else
   $A \leftarrow \operatorname{argmax}_{a \in \text{attributes}} \text{IMPORTANCE}(a, \text{examples})$ 
  tree  $\leftarrow$  a new decision tree with root test  $A$ 
  for each value  $v_k$  of  $A$  do
    exs  $\leftarrow$  {  $e : e \in \text{examples}$  and  $e.A = v_k$  }
    subtree  $\leftarrow$  DECISION-TREE-LEARNING(exs, attributes -  $A$ , examples)
    add a branch to tree with label ( $A = v_k$ ) and subtree subtree
  return tree

```

2.pg. 765 Excercise18.17 [10 points] K-NN

Suppose a 7-nearest-neighbors regression search returns $\{7, 6, 8, 4, 7, 11, 100\}$ as the 7 nearest y values for a given x value. What is the value of \hat{y} that minimizes the L1 loss function on this data? There is a common name in statistics for this value as a function of the y values; what is it? Answer the same two questions for the L2 loss function.

* The exercise deals with finding estimate of y given 7 nearest values to y using L1 and L2 loss function. What are the common names for the estimate of the y values?

3. K-NN [5 points]

Consider the training set below.

Ex #	A	B	C	D	Output
1	1	1	0	1	1
2	1	1	1	1	0
3	0	0	1	0	1
4	0	1	0	1	0
5	1	1	0	0	1
6	1	0	1	1	0

- A,B,C, and D are features

Assume you wish to use the K-nearest neighbor algorithm on this dataset and set aside the last two examples as a tuning set.

Would you prefer K=1 or K=3? Justify your answer.

4. pg. 766 Excercise 18.19 [10 points] SVM

Construct a support vector machine that computes the XOR function. Use values of +1 and -1 (instead of 1 and 0) for both inputs and outputs, so that an example looks like $([-1, 1], 1)$ or $([-1,-1],-1)$. Map the input $[x_1, x_2]$ into a space consisting of x_1 and $x_1 x_2$.

Draw the four input points in this space, and the maximal margin separator. What is the margin? Now draw the separating line back in the original Euclidean input space.

* Please start by drawing the 2D plain and moving it up to 3D in the feature space.

X1	X2	Y (XOR)
1	1	-1
1	-1	1
-1	1	1
-1	-1	-1

5. pg. 766 Excercise 18.21 [10 points] ANN

Construct by hand a neural network that computes the XOR function of two inputs. Make sure to specify what sort of units you are using.

* Remember, we were not able to construct single layer perceptron to do XOR operation. Start by think about what combination of logical operations is needed to compute XOR.

5. ANN [8 points]

Suppose you want to construct neural network that determines who gets to buy dinner that based on six votes. The voter (input) will either vote A or B.

- [5points] Draw the neural network for the system.
- [3 points] Assuming that each voter has equal privilege, what would the weight of each edges be?