

# CST8390 - Lab 3

## K Nearest Neighbor (kNN)

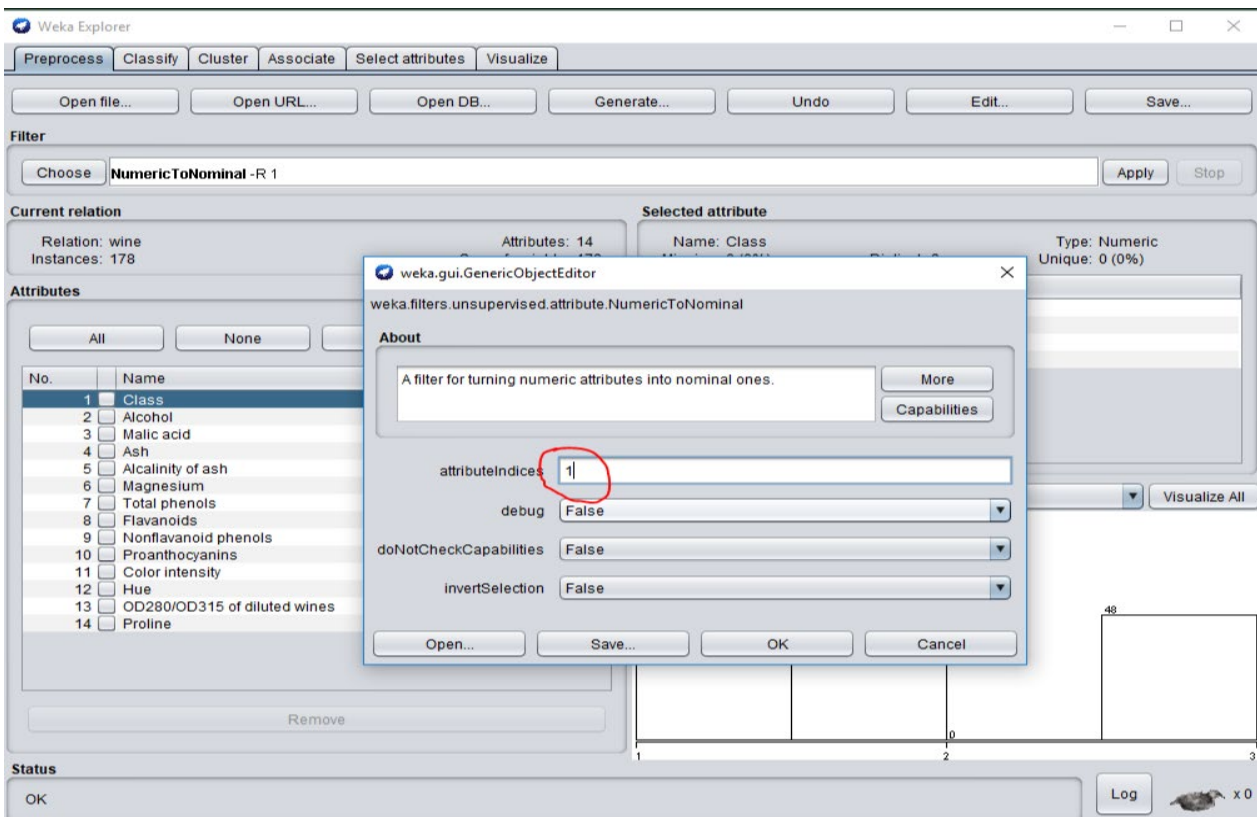
**Due Date:** Week 4 in corresponding lab sessions

### Introduction

The goal of this lab is to perform classification on wine dataset using kNN.

### Steps:

1. Create a word document named Lab3\_Answers.doc.
2. Get “Wine” dataset from <https://archive.ics.uci.edu/ml/datasets.php> (this link will show you a list of datasets, you need to search for wine dataset in this page) and save the content of the file as a csv file. (data is in wine.data and info is in data.names). Add attribute names as the first row in the csv file. (For every row, first value is the class, remaining values are various attributes)
3. Explore and learn about the relevance of various attributes of the dataset.
4. Load the file to Weka.
5. Check how various attributes are converted in Weka. Class is considered as numeric instead of nominal. Apply filter NumericToNominal to convert class datatype to nominal. When you apply filter, you need to specify the index of the attribute you need to apply the filter. Make sure that you are converting only the attributes to be converted. If you didn't specify attribute index, by default, all numeric attributes will be converted.



Now, you should see like this:

**Weka Explorer**

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

**Filter**

Choose **NumericToNominal -R 1** Apply Stop

**Current relation**

Relation: wine-weka.filters.unsupervised.a... Attributes: 14  
Instances: 178 Sum of weights: 178

**Attributes**

All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> Class
2	<input type="checkbox"/> Alcohol
3	<input type="checkbox"/> Malic acid
4	<input type="checkbox"/> Ash
5	<input type="checkbox"/> Alkalinity of ash
6	<input type="checkbox"/> Magnesium
7	<input type="checkbox"/> Total phenols
8	<input type="checkbox"/> Flavanoids
9	<input type="checkbox"/> Nonflavanoid phenols
10	<input type="checkbox"/> Proanthocyanins
11	<input type="checkbox"/> Color intensity
12	<input type="checkbox"/> Hue
13	<input type="checkbox"/> OD280/OD315 of diluted wines

Remove

**Selected attribute**

Name: Class  
Missing: 0 (0%)  
Distinct: 3  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	1	59	59.0
2	2	71	71.0
3	3	48	48.0

Class: Class (Nom) Visualize All

**Status**

OK Log x 0

- Now, we need to perform classification using kNN method. For that, click on “Classify” tab. For that, choose IBk which is Instance Based k Nearest Neighbors from Lazy in the tree view.

**Weka Explorer**

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

**Classifier**

weka

- classifiers
- bayes
- functions
- lazy
  - IBk**
  - KStar
  - LWL
- meta
- misc
- rules
- trees

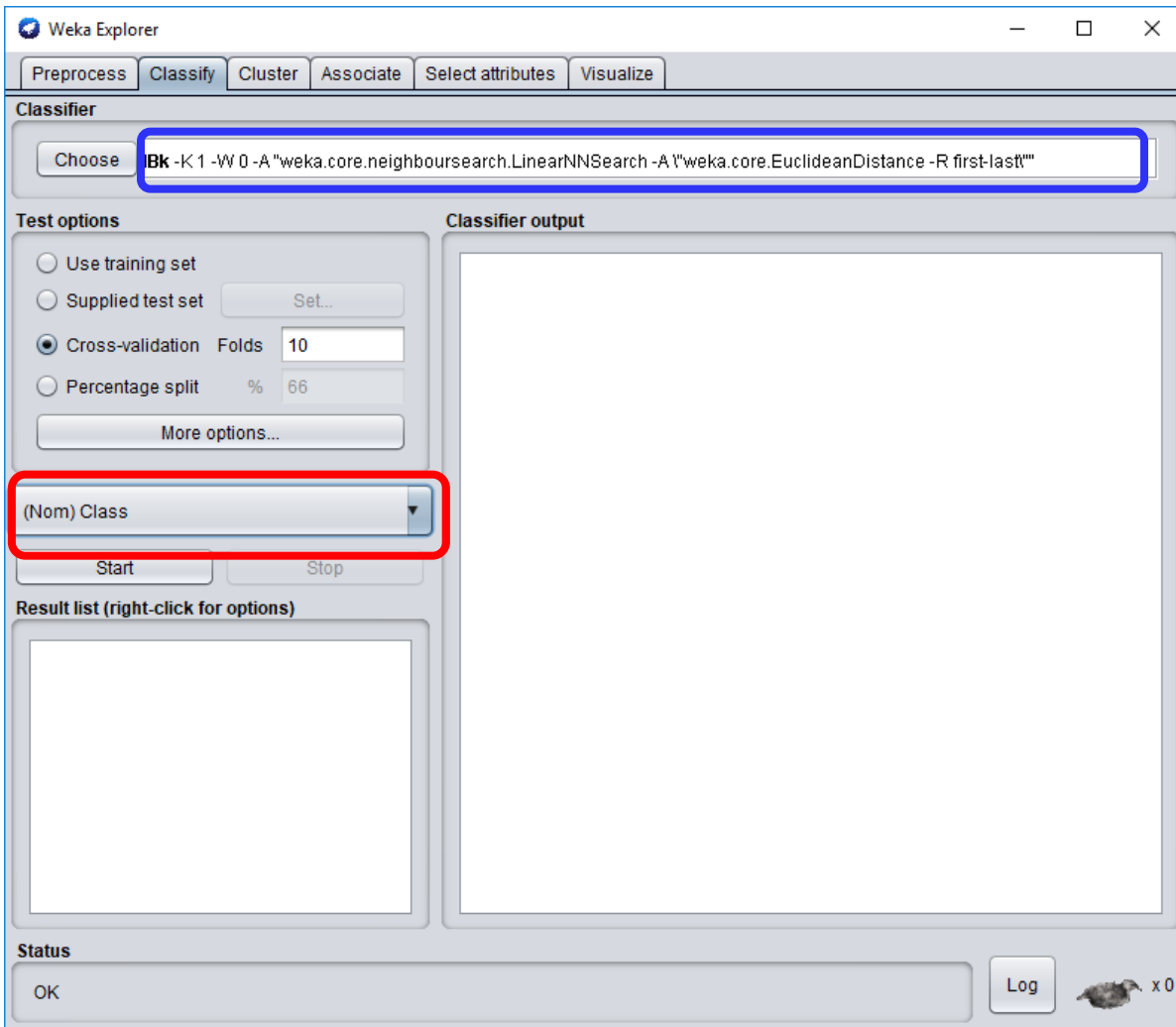
Close

**Status**

OK Log x 0

7. As mentioned earlier, our first attribute is the class label. We need to set that now in the classify panel. (Marked in red below)

IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch ...." This is the parameter list for the algorithm (Marked in blue). Click on this text to set the value of k. Set k as **1**. Close the window. Now, set the cross-validation to 10 Folds if it's not already there. Now click "Start" to run the algorithm.



8. There should be a lot of text in the right-hand side of the window with the results of the algorithm. Find the line that says "Correctly classified instances".

- What is the **percentage** of correctly classified items?
- What are the True Positive (TP) rates of **each** class?
- Look at the confusion matrix, which class is incorrectly classified?

Write these answers in the Answer document.

9. Now click on the “Choose” button to modify the number of neighbours that are used in the kNN search to **3**.
- d. What is the **percentage** of correctly classified instances? \_\_\_\_\_
  - e. What are the True Positive (TP) rates of **each** class? \_\_\_\_\_
  - f. Look at the confusion matrix, which classes are incorrectly classified? \_\_\_\_\_

Write these answers in the Answer document.

10. Run the algorithm several times, always increasing the value of N by two, and always an odd number: 5, 7, 9, 11, 13. Each of your tests will be in the window of the lower left. Fill in the following table.

g.

K	Percentage of correctly classified instances	Number of instances misclassified in each class
5		1: 2: 3:
7		1: 2: 3:
9		1: 2: 3:
11		1: 2: 3:
13		1: 2: 3:
15		1: 2: 3:

Include this table in the Answer document.

11. Repeat step 9 with “Percentage Split” of 70. Fill in the following table.

K	Percentage of correctly classified instances	Number of instances misclassified in each class
1		1: 2: 3:
3		1: 2: 3:
5		1: 2: 3:
7		1: 2: 3:
9		1: 2: 3:
11		1: 2: 3:
13		1: 2: 3:

Include this table in the Answer document.

You will have 15 executions on the left panel (8 with cross validation and 7 with percentage split).

In order to get credit for this lab,

1. You should be ready to show these 15 executions in Weka during demo.
2. Your answer document should have answers for questions 8,9,10, and 11.
3. Upload the answer document to Brightspace.

This lab has **5 marks** so ensure that you have all your answers filled in.

**Both demo during lab hours and submission in Brightspace are required to get credits for the lab.**