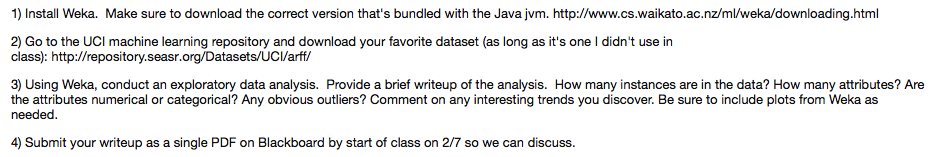
**Data Mining\_Assignment 1**

**Spring 2017**

**Kaiqin Huang**



**tic-tac-toe.arff**

**How many instances are in the data?**

958 instances.

**How many attributes?**

10 attributes. They are:

1. Top-left-square
2. Top-middle-square
3. Top-right-square
4. Middle-left-square
5. Middle-middle-square
6. Middle-right-square
7. Bottom-left-square
8. Bottom-middle-square
9. Bottom-right-square
10. Class

**Are the attributes numerical or categorical?**

All the attributes are categorical.

The first nine attributes have three distinct, b, o, and x.

The tenth attribute “Class” has values either negative or positive.

**Any obvious outliers?**

Not much, as far as I see. The proportion of “b” in the middle-middle-square is relatively smaller than other squares. Also in the middle-middle-square, proportion of “negative” is greater than “positive” for “o” option, and this is the only case where “negative” exceeds “positive”.

No missing data for all attributes.

For datasets with numerical attributes, it should be much more common to have outliers.

**Comment on any interesting trends you discover.**

“b” always has the smallest proportion and “x” always has the largest for any square. For attribute “Class”, “positive” surpasses “negative”.

I never played the tic tac toe game. Maybe I can give more analytical and deeper comments after I really play it.

**Be sure to include plots from Weka as needed.**

