M6L1 Homework Assignment (due 29th Oct)

This homework assignment focuses on k-Nearest Neighbors. You will provide a written analysis based on the following information:

- First, go to to the UC Irvine Machine Learning Repository and find a dataset for supervised classification. Every student MUST use a different dataset and it can be the same dataset you used for the unsupervised clustering as long as the data has some labeled data.
- Next, classify your data using k-Nearest Neighbors.
- Finally, answer the following questions:
    1. Does the k for kNN make a difference? Try for a range of values of k.
    2. Does scaling, normalization or leaving the data unscaled make a difference for kNN? Why or Why not?

As a reminder, please provide a written analysis/report as an .Rmd file. Note: This is a graded assignment due by Sunday at 11:59 pm.

M6L2 Homework Assignment (29 Oct)

This homework assignment focuses on Decision Trees. You will provide a written analysis based on the following information:

- First, go to to the UC Irvine Machine Learning Repository and find a dataset for supervised classification. Every student MUST use a different dataset and it can be the same dataset you used for the unsupervised clustering as long as the data has some labeled data.
- Next, generate a Decision Tree with your data. You can use any method/package you wish.
- Finally, answer the following questions:
    1. Does the size of the data set make a difference?
    2. Do the rules make sense? If so why did the algorithm generate good rules? If not, why not?
    3. Does scaling, normalization or leaving the data unscaled make a difference?

As a reminder, please provide a written analysis/report as an .Rmd file. Note: This is a graded assignment due by Sunday at 11:59 pm.

M6L3 Homework Assignment (due 5th Nov)

This homework assignment focuses on Support Vector Machines (SVMs). You will provide a written analysis based on the following information:

- First, go to to the UC Irvine Machine Learning Repository and find a dataset for supervised classification. Every student MUST use a different dataset and it can be the same dataset you used for the unsupervised clustering as long as the data has some labeled data.
- Next, classify your data using Support Vector Machines. You can use any method/package for SVMs.
- Finally, answer the following questions:
    1. How well does the classifier perform?
    2. Try different kernels. How do they effect its performce?
    3. What might improve its performce?

As a reminder, please provide a written analysis/report as an .Rmd file. Note: This is a graded assignment due by Sunday at 11:59 pm.

M6L4 Homework Assignment (due 5th Nov)

This homework assignment focuses on Linear Discriminant Analysis (LDA). You will provide a written analysis based on the following information:

- First, go to to the UC Irvine Machine Learning Repository and find a dataset for supervised classification. Every student MUST use a different dataset and it can be the same dataset you used for the unsupervised clustering as long as the data has some labeled data.
- Next, classify your data using LDA.
- Finally, answer the following questions:
    1. Does the number of predictor variables for LDA make a difference? Try for a range of models using differing numbers of predictor variables.
    2. What determines the number of linear discriminants in LDA.
    3. Does scaling, normalization or leaving the data unscaled make a difference for LDA?

As a reminder, please provide a written analysis/report as an .Rmd file. Note: This is a graded assignment due by Sunday the at 11:59 pm.