Practical Data Science
(Model Evaluation & Tuning)

Solve the following problems.

Problem1: Classification Accuracy

a. Suppose you are working on stock market prediction. You would like to predict whether or not a certain company will declare bankruptcy within the next 7 days (by training on data of similar companies that had previously been at risk of bankruptcy). Would you treat this as a classification or a regression problem?

b. Calculate the Accuracy/Error for a classification problem with the following response and class estimate:

y	$\hat{y}$
0	0
1	1
0	1
1	0
1	1
0	1
0	0

Problem2: Classification Accuracy using K-fold cross-validation

Below is a data set with response *y* and class predictions ^ *y*. Perform K-fold cross-validation with *K* = 3 by computing CV accuracy/error..

a. When we let our first fold be predicted using folds 2 and 3 as training data we get:

y	$\hat{y}$
1	1
1	1
1	0
0	0
1	1
0	0

b. When we let our second fold be predicted using folds 1 and 3 as training data we get:

Practical Data Science
(Model Evaluation & Tuning)

y	$\hat{y}$
0	1
0	0
1	1
1	0
1	1
1	1
0	0

c. When we let our third fold be predicted using folds 1 and 2 as training data we get:

y	$\hat{y}$
0	1
0	0
1	1
1	1
0	0
1	1
0	0

Problem3: A Second attempt at Predictive Analytics Problems

Go through the following problem of kaggle:
https://www.kaggle.com/c/ghouls-goblins-and-ghosts-boo/data

Do the following tasks:

a. Apply decision tree ML algorithm to discover the pattern automatically. Use default parameters for decision tree algorithm. Find out the CV error and train error. Do you see overfitting with the learned model? Find out the test accuracy after submitting to kaggle.
b. Apply model tuning of decision tree algorithm with different values for its parameters and build the model in the ideal range and findout CV error. ? Find out the test accuracy after submitting to kaggle.