Conclusion of Part 3 - Classification

Section 19, Lecture 132

In this Part 3 you learned about 7 classification models. Like for Part 2 - Regression, that's quite a lot so you might be asking yourself the same questions as before:

1. What are the pros and cons of each model ?
2. How do I know which model to choose for my problem ?
3. How can I improve each of these models ?

Again, let's answer each of these questions one by one:

1. What are the pros and cons of each model ?

Please find here (http://www.superdatascience.com/wp-content/uploads/2017/02/Classification-Pros-Cons.pdf) a cheat-sheet that gives you all the pros and the cons of each classification model.

2. How do I know which model to choose for my problem ?

Same as for regression models, you first need to figure out whether your problem is linear or non linear. You will learn how to do that in Part 10 - Model Selection. Then:

If your problem is linear, you should go for Logistic Regression or SVM.

If your problem is non linear, you should go for K-NN, Naive Bayes, Decision Tree or Random Forest.

Then which one should you choose in each case ? You will learn that in Part 10 - Model Selection with k-Fold Cross Validation.

Then from a business point of view, you would rather use:

- Logistic Regression or Naive Bayes when you want to rank your predictions by their probability. For example if you want to rank your customers from the highest probability that they buy a certain