Lecture 5: *"Random Forests"*

Book Chapter

Please read the *Chapter 7* **"Random Forests"** without the Regression part in "Gradient Boosting" and answer the following Questions.

Keep in mind: If you answer these questions and write a detailed summary, you won't need to read these chapters again while preparing for the exam

Video Nugget

This is a nice series about Random Forests. There is also a video about AdaBoost!

Questions

1. What is the idea behind *ensemble learning*?
2. Explain how ensemble learning work. What is the difference between a hard voting classifier and a soft voting classifier?
3. What can improve the ensemble's accuracy?
4. What is the difference between *bagging* and *pasting*?
5. How does *out of bag* evaluation work?
6. Is it beneficial to have a higher randomness in our forest? Explain why or why not.
7. What other way to bring randomness into the ensemble exist than varying the subsets which are handled from each model?

    - Random Forest, Extra-Trees and Feature importance in the meeting

8. What is the basic idea behind Boosting? What is a drawback?
9. How does AdaBoost work?

    - Gradient Boosting for classification in the meeting, for regression later

10. What is the idea behind stacking?

Homework Assignment

Please work on the exercises given in Ensemble_Learning.zip .