

IST565 Data Mining

HW7: SVMs, kNN, and Random Forest for handwriting recognition

In this homework, you will use SVMs, kNN, and Random Forest algorithms for handwriting recognition, and compare their performance with the naïve Bayes and decision tree models you built in previous week.

Steps:

1. Describe data pre-processing steps and the chosen evaluation method and measure(s)
2. Use the train set to build kNN, SVM, and Random Forest models. Submit these models' prediction results to Kaggle. Report test performance, compare them, and use the theoretic knowledge to explain whether the algorithm performance difference makes sense or not.
3. Write a report to describe what you did, including the data preparation, transformation, algorithm tuning, the generated models and their performance. In the end, summarize which model works the best and why.
4. If you use Weka to do the experiment, write your report in Microsoft Word. Up to 8 pages. NO PDF PLEASE.
5. If you use R, use R Markdown tool to prepare your document and output to Word. **Don't print out excessively large amount of output**, such as prediction results, or the entire dataset.

Grading rubrics:

1. Are the models constructed correctly?
2. Is the result analysis conclusion convincing?
3. Is sufficient details provided for others to repeat the analysis?
4. Does the analysis include irrelevant content?
5. Successful submission to Kaggle?