

Friday: Independent Project - Week 8

Agenda

1. **Instructions**
 - Learning Outcomes
 - Deliverables
2. **Assessment**
3. **Submission & Evaluation**

Instructions

Assessment

Overview

Nairobi Hospital conducted a clinical camp to test for hypothyroidism. The data collected focused on Thyroid patients. Use the healthcare datasets provided to accomplish the following:

Build a model that determines whether or not the patient's symptoms indicate that the patient has hypothyroid.

Part 1: Decision trees:

- For this section, you should build a model that makes the above prediction. You should not use individual decision trees, rather you should use at least 2 out of the 3 advanced models we have studied: Random forests, Ada boosted trees, and gradient boosted trees.
- Try and optimize each of the 2 models, making sure to document how you've set up your hyperparameters.
- Identify which of the 2 models you trust most, and use your model to determine **which features are most impactful in influencing the prediction**
- Note that with decision trees, you don't need to do a lot of data cleaning. This will be very different with SVM.

Part 2: SVM:

- In this section, you may be required to clean the data a little bit so as to make sense of the features.
- Document what transformation you've done on the data.
- Apply Polynomial, linear and rbf kernel function to build your SVM model and then evaluate their performance and pick the kernel that performs the best. Remember to tune your parameters to improve the performance of your model. **To make your life easier, make**

sure to visualize the models you've created. Use any two features to build the models for this step.

Hint: You may want to use decision trees to give you the most preferable features you can use. but also keep in mind that those features might not be suitable for SVM. It might be a good idea to graph them first.

- After getting your best performing kernel, use this kernel together with your tuned parameters and repeat the prediction but this time using additional features. Compare the model you've just created with the 2-features version.

Dataset

Dataset Files

Source [\[Link\]](http://bit.ly/hypothyroid_data) [_\(http://bit.ly/hypothyroid_data\)](http://bit.ly/hypothyroid_data).

Dataset Columns

- Age
- Sex
- on_thyroxine
- query_on_thyroxine
- on_antithyroid_medicationthyroid_surgery
- query_hypothyroid
- query_hyperthyroid
- pregnant
- sick
- tumor
- lithium
- goitre
- TSH_measured
- TSH
- T3_measured
- T3
- TT4_measured
- TT4

Submission & Evaluation

The submission to this week's Independent Project should be made here [\[Link\]](https://moringaschool.instructure.com/courses/624/assignments/9923) [_\(https://moringaschool.instructure.com/courses/624/assignments/9923\)](https://moringaschool.instructure.com/courses/624/assignments/9923)].

Note:

- The deadline for this assessment is 6.00pm on the day of the assessment.
- Late submissions will not be assessed.

- Do not seek to copy someone else's work while working on this Independent project. You deny yourself an opportunity to learn whenever you resolve to plagiarism.