Yu Chang Ou

CSC321: Data Mining and Machine Learning

Final Project

**Dataset**: Thyroid Disease Data Set

**Link**: https://archive.ics.uci.edu/ml/datasets/Thyroid+Disease

**Number of Instances**: 7200

**Number of Attributes**: 25

Category/Class Measurement

1. age: continuous.
2. sex: M, F.
3. on thyroxine: f, t.
4. query on thyroxine: f, t.
5. on antithyroid medication: f, t.
6. sick: f, t.
7. pregnant: f, t.
8. thyroid surgery: f, t.
9. query hypothyroid: f, t.
10. query hyperthyroid: f, t.
11. lithium: f, t.
12. goitre: f, t.
13. tumor: f, t.
14. TSH measured: f, t.
15. TSH: continuous.
16. T3 measured: f, t.
17. T3: continuous.
18. TT4 measured: f, t.
19. TT4: continuous.
20. T4U measured: f, t.
21. T4U: continuous.
22. FTI measured: f, t.
23. FTI: continuous.
24. TBG measured: f, t.
25. TBG: continuous.

**Attribute Files**: hypothyroid, thyroid0387, sick-euthyroid, new-thyroid, sick, dis, ann-train

**Purpose**: The purpose of this data set is to classify a patient of whether they have *Hyperthyroidism* (Excessive Production of Thyroid Hormone), *Hypothyroidism* (Underproduction of Thyroid Hormone), Euthyroid/*Normal* (Normal levels of thyroid hormone production, or problem is not caused by imbalance of thyroid hormone generated by the thyroid gland.) Given a patient’s medical history and blood test, we will utilize the following 25 classes to use machine learning to determine the state of the medical anomaly and classify hyper, hypo, or normal. We will continuously work with the data to further determine the 5 most important factors which attributes to the class of the illness.