Machine Learning – COMP3032

Tutorial and Lab Practice 3 – Week 4

This lab pactice focuses on the concept of supervised learning: classification.

Tutorial

1. Review the terminology introduced and concepts taught in Lecture 3.
2. Review related concepts of linear algebra.
3. What is logistic regression? What is it used for?
4. Compare linear regression and logistic regression.
5. What is the general form of log loss function? Does it have a closed form solution? Is it convex?

Lab Practice

1. Download, open and run the program tut03-iris.py. Read and understand the program.
2. Revise tut03-iris.py:
    1) Build a classifier to detect the Iris *setosa type* based only on the petal width feature.
    2) From the plot, what is the range of petal width for Iris *setosa*?
    3) Use the classifier to predict if the flowers are of the *setosa* type if their petal widths are 0.3, 0.75, and 1.5 cm respectively.
    4) Change the above predictions to probabilities.
3. The files students1.csv and students2.csv, available on vUWS, contain examples about subject results, pass(1)/fail ()0), versus average study time per week.
    1) Use scikit-learn to compute a logistic regression model from students1.csv for the classification.
    2) Use the model to classify the (training) instances in students1.csv. How many of the examples are misclassified? What are the precision and recall values?
    3) Use the model to classify the instances in students2.csv. How many of the instances are misclassfied in this case? What are the precision and recall values?