University of the Witwatersrand, Johannesburg
School of Computer Science and Applied Mathematics

Laboratory Assessment: Machine Learning (COMS 3007)

Phumlani Nhlanganiso Khoza and Benjamin Rosman

Due Date: 26 February 2018

In this laboratory assessment we expect the following deliverables from each student:

- Implementations of algorithms.
- Short reports containing experiment results.

Group discussions are encouraged, but it is expected that implementation is largely done independently. The submission deadline is 14:00 on 26 February 2018.

1 Optimization: Direct Approach

1.1. Outline one approach to directly obtaining optimal model parameter values when given a data set, a model, and an objective function that is differentiable in its parameters.

1.2. Derive the *normal equations* assuming that the regression model (i.e. hypothesis) is:

(a) Linear
(b) Exponential

1.3. Generate noisy data that are consistent with the hypotheses in Question 1.2, and implement a programme that performs the parameter fitting procedure.

(a) What is the effect of changing the functional form of the noise model?
(b) What is the effect of changing the parameter values of the noise model?

2 Optimization: Iterative Approach

2.1. Under what conditions do we need to use iterative optimization techniques?

(a) Why do gradients provide a useful approach?
(b) Apart from gradients, what other approach could we possibly use (even if not very efficient)?

2.2. Derive the single variable Newton-Raphson method iterative equation $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$.

(a) Why is this algorithm useful in a machine learning context?
(b) In a machine learning context, what are the interpretations of: $x_i, f(x_i), f'(x_i)$?

2.3. Extend the derivation in Question 2.2 to an nD space.

2.4. Implement the multi-variable Newton-Raphson algorithm.

2.5. Using the algorithm implemented in Question 2.4, report the *performance* of the method. **Hint:** design an experiment to demonstrate the different aspects of *performance*.

2.6. *Investigating other ideas:

(a) What insights can be incorporated into the algorithm to assist with faster convergence, problems would be entailed by the inclusion of those insights?

(b) Discuss topological properties of the model would lead to poor performance when using the Newton-Raphson method. **Hint:** what could happen if we imposed constraints on the allowed parameter values?

(c) Discuss geometrical properties of the model that would lead to poor performance when using the Newton-Raphson method. **Hint:** what would happen if the surface on which the parameters lie were to be very wrinkly?