School of Computing and Information Systems
The University of Melbourne
COMP30027 Machine Learning (Semester 1, 2021)

Workshop: Week 7

1. What is **gradient descent**? Why is it important?

2. What is Linear Regression? In what circumstances is it desirable, and in what circumstances it is undesirable?

3. Recall that the update rule for Gradient Descent with respect to Mean Squared Error (MSE) is as follows:

$$\beta_k^{j+1} = \beta_k^j + \frac{2\alpha}{N}\sum_{i=1}^{N} x_{ik}(y_i - \hat{y}_i^j)$$

   Build a Linear Regression model, using the following instances:

x	y
1	1
2	2
2	3

4. What is **Logistic Regression**? What is "logistic"? What are we "regressing"?

5. [OPTIONAL] What is the relation between "odds" and "probability"?

6. Build a logistic regression classifier, which uses the counts of selected words in the news articles to predict the class of the news article (fruit vs. computer). Assume the weights of the 4 features (and the bias β_0) is $\widehat{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4] = [0.2, 0.3, -2.2, 3.3, -0.2]$.

ID	*apple*	*ibm*	*lemon*	*sun*	CLASS	
TRAINING INSTANCES						
A	1	0	1	5	1	FRUIT
B	1	0	1	2	1	FRUIT
C	2	0	0	1	1	FRUIT
D	2	2	0	0	0	COMPUTER
E	1	2	1	7	0	COMPUTER
TEST INSTANCES						
T	1	2	1	5	0	COMPUTER

   a) Explain the intuition behind the model parameters, and their meaning in relation to the features

   b) Predict the test label.

7. For the model created in question 6, compute a single gradient ascent update for parameter β_1 given the training instances given above. Recall that for each feature j, we compute its weight update as

$$\beta_j \leftarrow \beta_j + \alpha \frac{\partial \mathcal{L}(\beta)}{\partial \beta_j}$$

   Summing over all training instances i. We will compute the update for β_j assuming the current parameters as specified above, and a learning rate $\alpha = 0.1$.