Authors: Wojciech Ptaś (7042843), Mhd Jawad Al Rahwanji (7038980)

Problem 1

1. Logistic regression provides probabilities and is applied for classification (predicting categorical outputs) based on continuous and/or discrete inputs.
   When using linear regression for classification, some problems that we may face:
   - Inability to generalize to more than 2 classes: Although linear regression could work for binary classification (2 classes), however, it will fail in the case of multiple classes. If the response were binary, the classes can be factorized to 0 and 1 so that linear regression would predict them as if they were continuous values. Whereas in a multi-class scenario, factorization forces ordinality which would mislead the model. Unless we can order the classes, any factorization would be ambiguous resulting in different models that should be the same. Additionally, the masking problem can occur. It is an outcome of linear regression rigidity, that is, when we have more than 2 classes, a line can divide the space into 2 zones at best. The remaining classes would be masked by one of the other 2 classes rendering it indistinguishable.
   - Out of range [0, 1]: When fitting a linear regression to a binary response, the line will predict values above and below the y=[0, 1] range for big/small enough input values.
   - Outliers/High leverage points have a worse effect: For the sake of simplicity, in the case of binary classification using a single continuous input, our intuition would suggest classifying based on finding a threshold on the x-axis which maximizes correct predictions ($x \leq$ threshold $\Rightarrow$ class A and vice versa). But due to the rigidity of linear regression, each value of x gets a corresponding y value. The problem lies here, an extreme/outlier value (i.e. a large positive value) on the x-axis would tilt the fitted line causing many previously correctly predicted points to be misclassified.
   - Inappropriate performance metric: Since the outputs are categorical and linear regression uses least squares to fit the data, the R^2 metric, which relies on RSS, wouldn't be of value in model evaluation and comparison.