ECE 411 Homework 7 (Fall 2023)
Instructor: Dr. Chau-Wai Wong
Material Covered: Multiple Linear Regression, k-NN

Problem 1 (60 points) Watch this 70-min *supplemental lecture video* on multiple linear regression and answer the following questions:

a) Use the example on the coins or number of tackles to explain why it is imprecise to interpret β_j as the average effect on Y of a one-unit increase in X_j.

b) Focus on the 4th row of the first table on slide 20 "Results for advertising data." What hypothesis testing problem does this row correspond to? Explicitly give the mathematical form of H_0. Explain what a p-value of 0.8599 implies.

c) Prove that for p candidate predictors, there are $2^{p+1} - 1$ possible regression problems.

d) Based on what you have learned from "Qualitative predictors with more than two levels," when there are 4 categories, how many dummy variables/predictors need to be created? What coding pattern does each category have?

e) Explain in mathematical expression how predictors "radio" and "TV" are interacting with each other.

f) Given a set of data points that may be generated by a 5th or 6th-order polynomial regression function and an additive noise, how would you determine the order of the regression function by using a hypothesis testing package in R or Python?

Problem 2 (20 points) [Interaction, Nonlinear Transformations, Qualitative Predictors] Read the "Nonlinear Relationship" subsection on p.90 of ISLR and complete *ISLR-3.6.4–6* and write a concise report.

For Python users, please download *Boston.csv data* and *Carseats.csv data* and follow the text book's instructions while referring to *the "equivalence" Python codes* of ISLR-3.6.4–6.

Problem 3 (Bonus, 30 points) [k-Nearest Neighbors] (Try this problem after 10/23 lecture)

a) (20 points) Complete *ISLR-2.4.7*. Repeat (a)–(c) for $(X_1, X_2, X_3) \in \{(1, 2, 3), (1, -1, 1)\}$.

b) (10 points) Using a programming language of your choice, refactor your code into a function named `MyKnn` with the following input and output variables. We have shown below examples in Python and Matlab, but you may also use R.

Python:

```
def MyKnn(x1, x2, x3, k):
  ...
  return Y
```

Matlab:

```
function Y = MyKnn(x1, x2, x3, k)
...
end
```

The file containing the function should be named `MyKnn` with extension `.py`, `.m`, or `.r` and appended to the homework submission in plaintext. The performance of `MyKnn` will be manually assessed, and bonus points will be given solely on the percentage of correct

classifications using test data. You can assume that when the function is evaluated, the input variables `x1`, `x2`, `x3` will be any value in $\mathbb{R}$, `k` will be less than 6, and the return value being checked against will be either `"Red"` or `"Green"`.

Problem 4 (Bonus, 10 points) Complete ISLR-3.7.9. (Doing this problem may help resolve challenges that you may encounter in topic 1 "North Carolina road safety analysis and prediction" of the term project.)