Biostatistics 615 Learning Exercise #3 (10 pts)

Due by September 17th 2024 (Tuesday) 11:59pm. Use Gradescope (via Canvas) to submit an R file.

- Your submission should only contain one `R` file named `predProb.R` that contains a function named `predProb(x)`.
- Your code will be evaluated in Gradescope using 10 different test cases using an automated script. Full credit will be given if your code passes all test cases.
- You are allowed to submit multiple times before the deadline, but only the last submission will be graded. Automated feedback will be provided for each submission.
- You need to implement the function to work with arbitrary (valid) input values beyond the 10 cases tested. If you tweak your implementation so that your functions works specifically for the test cases, you will not receive any credit.
- Implement your function as efficient as you can. If your program does not finish after running for 1.0 seconds, you will lose the points for those test cases. Note that the official solution finishes within 0.2 seconds for any test case, so this should be a reasonable time limit.

Problem 1 - `predProb.R` (10 pts)

Write an R function `predProb.R` that contains a function `predProb(x)`, which computes the predictive probability of a fitted five-category logistic regression model for the classification problem:

$$\widehat{p}_k := \widehat{\Pr}(Y = k \mid X_1, \ldots X_p) \propto \exp\left(\widehat{\beta}_{k,0} + \sum_{j=1}^{p} \widehat{\beta}_{k,j} X_j\right), \quad \text{for } k = 1, \ldots, 5,$$

where the estimated coefficient $\widehat{\beta}_{k,j} = 2^{|k-j|}$ for $j = 1, \ldots, p$ and $\widehat{\beta}_{k,0} = 2^{-k}$ for $k = 1, \ldots, 5$. The predictor $X_j \in [0, 1]$ for $j = 1, \ldots, p$ is given as an input argument `x` as a vector of length p. The positive integer p is the number of predictors in the model. It is determined by the number of X_j's that are entered into the argument list of the program. The notation "$a_k \propto b_k$" means that a_k/b_k is a constant for $k = 1, \ldots, 5$.

You may assume that all input arguments are numeric values. Note that your `R` script file must be named as `predProb.R`. The input argument of the function is a vector of X_j's and the output returns $\widehat{p}_k$, for $k = 1, \ldots, 5$ as a vector of length 5.

For example, when `x = c(0, 0.1)`, the five probability values should be as follows (rounded to 8 digits):

```
0.23143717
0.16309102
0.15906429
0.18251063
0.26389689
```

The actual examples that will be used for grading will contain more complex examples, so be sure to retain maximum precision in your implementation, robustly across various types of input. At least 8 digit precision will be evaluated for correctness.

Note that you are NOT allowed to use any functions outside the `base` package in your implementation. Use `help(...)` to check whether a function belongs to the `base` package or not.

No error handling for malformed arguments is needed.