Assignment 3

CS 750/850 Machine Learning

- **Due**: February 17th at 11:59PM
- **Submisssion**: Turn in both a **PDF** and the **source code** on MyCourses
- **Questions**: Piazza and Office hours: *Marek*: Wed 1:30-3:00pm, Soheil: Mon 2-4pm, Xihong: Thu 1:30-3:30pm
- **Extra credit**: Especially good questions or helpful answers on Piazza regarding the assignment earn up to 5 points extra credit towards the assignment's grade.

Problem 1 [25%]

Suppose that I collected data for a group of machine learning students from last year. For each student, I have a feature X_1 = hours studied for the class every week, X_2 = overall GPA, and Y = whether the student receives an A. We fit a logistic regression model and produce estimated coefficients, $\hat{\beta}_0 = -6, \hat{\beta}_1 = -0.1, \hat{\beta}_2 = 1.0$.

1. Estimate the probability of getting an A for a student who studies for $40h$ and has an undergrad GPA of 2.0
2. By how much would the student in part 1 need to improve their GPA or adjust time studied to have a 90% chance of getting an A in the class? Is that likely?

Problem 2 [25%]

Consider a classification problem with two classes `T` (true) and `F` (false). Then, suppose that you have the following four prediction models:

- **T**: The classifier predicts `T` for each instance (always)
- **F**: The classifier predicts `F` for each instance (always)
- **C**: The classifier predicts the *correct* label always (100% accuracy)
- **W**: The classifier predicts the *wrong* label always (0% accuracy)

You also have a test set with 60% instances labeled `T` and 40% instances labeled `F`. Now, compute the following statistics for each one of your algorithms:

Statistic	Cls. T	Cls. F	Cls. C	Cls. W
recall				
true positive rate				
false positive rate				
true negative rate				
specificity				
precision				

Some of the rows above may be the same.

Problem O2 [30%]

This problem can be substituted for Problem 2 above, for up to 5 points extra credit. The better score from problems 2 and O2 will be considered.

Solve Exercise *3.4* in [Bishop, C. M. (2006). Pattern Recognition and Machine Learning].

Problem 3 [25%]

In this problem, you will derive the bias-variance decomposition of MSE as described in Eq. (2.7) in ISL. Let f be the true model, $\hat{f}$ be the estimated model. Consider fixed instance x_0 with the label $y_0 = f(x_0)$. For simplicity, assume that $\mathrm{Var}[\epsilon] = 0$, in which case the decomposition becomes:

$$\underbrace{\mathbb{E}\left[(y_0 - \hat{f}(x_0))^2\right]}_{\text{test MSE}} = \underbrace{\mathrm{Var}[\hat{f}(x_0)]}_{\text{Variance}} + \left(\underbrace{\mathbb{E}[f(x_0) - \hat{f}(x_0)]}_{\text{Bias}}\right)^2 .$$

Prove that this equality holds.

Hints:

1. You may find the following decomposition of variance helpful:

$$\mathrm{Var}[W] = \mathbb{E}\left[(W - \mathbb{E}[W])^2\right] = \mathbb{E}\left[W^2\right] - \mathbb{E}[W]^2$$

2. This link could be useful: https://en.wikipedia.org/wiki/Variance#Basic_properties

Problem 4 [25%]

Please help me. I wrote the following code that computes the MSE, bias, and variance for a test point.

```
set.seed(1984)
population <- data.frame(year=seq(1790,1970,10),pop=c(uspop))
population.train <- population[1:nrow(population) - 1,]
population.test <- population[nrow(population),]
E <- c()  # prediction errors of the different models
for(i in 1:10){
    pop.lm <- lm(pop ~ year, data = dplyr::sample_n(population.train, 8))
    e <- predict(pop.lm, population.test) - population.test$pop
    E <- c(E,e)
}
cat(glue::glue("MSE:         {mean(E^2)}\n",
               "Bias^2:      {mean(E)^2}\n",
               "Var:         {var(E)}\n",
               "Bias^2+Var:  {mean(E)^2 + var(E)}"))
```

```
## MSE:         2869.61343086216
## Bias^2:      2681.61281912074
## Var:         208.889568601581
## Bias^2+Var:  2890.50238772232
```

I expected that the MSE would be equal to Bias^2 + Variance, but that does not seem to be the case. The MSE is 2402.515 and Bias^2 + Variance is 2428.706. Was my assumption wrong or is there a bug in my code? Is it a problem that I am computing the expectation only over 10 trials?

Hint: If you are using Python and need help with this problem, please come to see me (Marek).