OM 352, Fall 2023 **HW 9** **Assigned: 23 Nov. 2023**
Due: 11:59 PM, 29 Nov. 2023

Instructions: Each assignment will include a PDF file (like this one) with the assignment questions and an Excel file, with an Answers sheet and any data or models that we provide you with. You must download both the PDF and Excel files. You have to enter your answers in the Answers sheet of the same Excel file you have downloaded, then save and upload the Excel file. You must upload the same Excel file you downloaded. Further instructions are provided in the Online Assignment Tools Guide (see **Assignments** in eClass).

Put your answers in the appropriate cells (salmon-colored cells) in the **Answers** sheet. Use paste special … values to for all numerical answers. The other cells in the Answers sheet are locked, which means you won't be able to enter values into those cells. Do not change the format of cells in the Answers sheet. Save your file with the appropriate name and in the appropriate format ("HW#_ID.xlsx").

Marking will be based on the answers in the Answers worksheet of the file you upload. We will only look at the rest of the file if there is an appeal (and even then, the answers in the Answers sheet take precedence.) If you wish to appeal a mark, then the file you uploaded must include your supporting work for each question. It is a good idea to make one worksheet for each question.

Total points: 18 points, of which 2 points are for following the submission instructions provided above.

Insurance Fraud Detection:

The GoFarm insurance company insures farmers in Alberta against damages caused by floods, hail, or tornados. When submitting a claim, a farmer needs to specify the amount claimed ($) and the area of their farm, in hectares. After receiving the claim, GoFarm might send an inspector to the farms to check whether the claim is valid or fraudulent.

Sending an inspector to a farm costs $3,000 per farm. If the claim is not fraudulent, then GoFarm pays the amount specified in the claim plus the $3,000 cost to send an inspector. If the claim is fraudulent, then the claim is not paid and the farmer is fined. The average amount of fine is $15,000 per fraudulent claim. So the net monetary consequence for the company is a revenue of $15,000-$3,000 = $12,000.

If GoFarm does not send an inspector to investigate a claim, then GoFarm simply pays the amount claimed.

GoFarm needs your help to predict which claims are most likely to be fraudulent and should be checked (by sending an inspector). You are given a training set (Worksheet "Training") of 40 claims from the previous year, showing the area in hectares, the amount claimed, and whether the claim was fraudulent. You are also given a test set (Worksheet "Test") that shows the area in hectares and the amount claimed for 35 claims made this year. The test set does not include information about whether the claims are fraudulent, because this information is still unknown.

In the following questions we will ask you to compute the total net cost incurred by GoFarm under different policies, where:

Total net cost = cost of paying claims + cost of sending inspectors
– money recovered through fines for fraudulent claims

1. (4 pts.) Using the data in the "Training" worksheet, compute the total net cost under the following four inspection policies:
    a. Do not send any inspectors—just pay all claims.
    b. Send inspectors to investigate all claims.
    c. Send inspectors to investigate only those claims involving more than 150 hectares.
    d. Send inspectors to investigate only those claims with amounts above $10,000.

2. (3 pts. for feasibility, 2 pts. for consistency, 1 pt. for optimality) Use the training set to find the line (c_1*Amount Claimed + c_2*Hectares = b) that best separates fraudulent from valid claims so as to minimize *the total number of misclassification errors*. Use M = 500 and use e = 0.1. Remember to include the normalization constraints ($-1 \le c_1, c_2 \le 1$). Report the following:
    a. The parameters c_1, c_2, and b
    b. The binary misclassification variable value for each claim in the training set
    c. The total number of misclassification errors in the training set, using the solution you found.
    d. Which claims in the test set should be inspected, using the solution you found.

3. (3 pts. for feasibility, 2 pts. for consistency, 1 pt. for optimality) Use the training set to find the line (c_1*Amount Claimed + c_2*Hectares = b) that best separates fraudulent from valid claims so as to minimizes *the total net cost in the training set*. Use M = 500 and use e = 0.1. Remember to include the normalization constraints ($-1 \le c_1, c_2 \le 1$). Report the following:
    a. The parameters c_1, c_2, and b
    b. The binary misclassification variable value for each claim in the training set
    c. The total net cost in the training set, using the policy you found.
    d. Which claims in the test set should be inspected, using the policy you found.

Hint 1: Use the WarmFleece example from the lectures as a starting point. Think about whether "Fraud" is more like "Cold-blooded" or like "Warm-blooded." When building your model, keep in mind that you will need to report the values of the binary misclassification variables by claim number (Column A in the "Training" worksheet).

Hint 2: The models in this assignment can be difficult for solver, because it has 40 binary variables. It may take longer than usual to find the optimal solution. This is normal and you do not need to do anything about it. You just have to wait a bit longer (depending on your computer, you might have to wait up to 5 minutes). Press the Esc key (you may have to hold it for a second or two) if you need to stop solver, then wait for a dialog box that allows you stop or continue. Your solver may also return the solution in a few seconds. That is normal too.

Hint 3: All the models in this assignment are linear and therefore you should use Simplex LP in the solver.