CMPSC 448: Machine Learning. Homework 2: Decision Theory
Due: January 27, 2020

1. INSTRUCTIONS

- You cannot look at anyone else's code.
- Fill in and upload hw2.py to gradescope.
- All code (except import statements) in hw2.py should be inside functions (importing hw2.py should not cause code to execute).
- Code must have comments and any constants should be stored in a variable defined near the top of your file.

2. DECISION THEORY

Question 1. *The Pennsylvania Chocolate Institute of Health has identified a new type of disease caused by a diet containing insufficient amount of chocolate. Treatment requires expensive imported chocolates, covered by most insurance plans. There are only two types of patients:*

  A. *Healthy chocolate eaters*
  B. *Patients suffering from low chocolate levels*

The insurance company has developed a test (machine learning model) which predicts p, the probability that the current patient is of type A. The company is experimenting with a cost model:

  - *The cost of treating a patient of type A is α*
  - *The cost of treating a patient of type B is some number **treatB***
  - *The cost of not treating a patient of type A is some number **notTreatA***
  - *The cost of not treating a patient of type B is some number **notTreatB***

You must write a program to help the insurance company explore the cost of making decisions using decision theory.

  hw2.py, fill in the function q1(p, treatB, notTreatA, notTreatB)*, where p is the predicted probability the patient is of type A. The output of this function should be the value of α for which the insurance company is indifferent between treating and not treating the patient.*

Question 2. *After years of research, great progress has been made in treating Chocolate Deficiency Syndrome (CDS). The procedure is for a patient to walk into a CDS clinic and take a simple test. The test is a "has CDS"/"does not have CDS" classifier that outputs p, the probability the patient has CDS. Based on the results, the doctor can either (1) treat the patient, (2) not treat the patient, (3) request a more definitive test (but expensive) test. The associated costs are as follows:*

  - *The cost of the additional test is some number **testCost***
  - *The cost of mistakenly treating a patient who does not have CDS is **treatHealthy***
  - *The cost of treating a patient with CDS is **treatCDS***
  - *The cost of not treating a healthy patient is **notTreatHealthy***
  - *The cost of mistakenly not treating a patient who has CDS is **notTreatCDS***

  In hw2.py, fill in the function q2(p, testCost, treatHealthy, treatCDS, notTreatHealthy, notTreatCDS) *so that it outputs 1 if the doctor should treat the patient, -1 if the doctor should not treat the patient, and 0 if the doctor should ask for more tests.*