Problem 2: Predicting patient risk in remote healthcare monitoring

A healthcare technology company is developing a system for monitoring patients with chronic conditions remotely. To provide timely interventions and prevent complications, the company is implementing a predictive model to assess the risk of a patient's condition worsening. Instead of relying solely on periodic in-person check-ups, which can be infrequent and costly, the company is introducing continuous monitoring through wearable devices. This system will collect real-time data on patients' vital signs and activity levels, allowing for early detection of potential health risks. Patients flagged as high-risk will receive immediate medical attention.

The system captures several key variables: `hr`, representing the patient's resting heart rate; `activity`, measuring daily physical activity; `sleep`, indicating the quality of sleep; and `spo2`, tracking blood oxygen saturation. The file `health.txt` contains the average system measurements for 60 monitored individuals over a week. Additionally, the variable `risk` indicates whether the patient required urgent medical intervention in the subsequent week (`yes` if medical intervention).

The objective is to build a classifier that can predict which patients are at high risk of needing urgent care based on the four health indicators mentioned above.

The cost of providing an urgent medical intervention is \$500 per patient, which is equivalent to the value of accurately identifying a patient who needs it. If a high-risk patient is not identified and subsequently experiences a serious health event, the estimated economic loss is \$100,000. The company estimates that 0.1% of patients monitored will be at significant risk. The classifier should be optimal in minimizing the expected cost of misclassification.

a) Which classification method would you use? Verify the underlying assumptions.

b) Provide the costs of misclassification, the prior probabilities and the priors adjusted for considering the cost of misclassification.

c) Build the corresponding classifier, providing its apparent error rate and an estimate of its actual error rate through leave-one-out cross-validation. What do you observe? Which characteristic of the classifier you are using is being highlighted here?

The company plans to use the classifier developed in (b) to monitor a cohort of 1000 patients during the next monitoring period.

d) How much should be budgeted for the cost of urgent medical interventions?

e) Compared to the previous strategy of providing interventions based on scheduled check-ups for all patients, does the new predictive monitoring approach lead to a lower expected cost of misclassification? How much would be saved for this cohort?

Upload your results here: `https://forms.office.com/e/LK0Zpk2gQm`