COMPSCI 4ML3 - Introduction to Machine Learning
Assignment 1 Solution

October 18, 2024

Alireza Daeijavad
McMaster University

---

1. [**80 points**] Programming Component. Follow this link to open the Google Colab environment and make a copy of the notebook. Include the answers/graphs/pictures/analyses of the four tasks in your final pdf report. Additionally, upload your modified Jupyter notebook that includes your code (as a separate ipynb file). (for clarifications, post your questions in the Team's Q/A channel).

   Solution: LINK

2. [**30 points**]"Driving high" is prohibited in the city, and the police have started using a tester that shows whether a driver is high on cannabis. The tester is a binary classifier (1 for positive result, and 0 for negative result) which is not accurate all the time:

   - If the driver is truly high, then the test will be positive with probability $1 - \beta_1$ and negative with probability β_1 (so the probability of wrong result is β_1 in this case)
   - If the driver is not high, then the test will be positive with probability β_2 and negative with probability $1 - \beta_2$ (so the probability of wrong result is β_2 in this case)

   Assume the probability of (a randomly selected driver from the population) being "truly high" is α.

   - **[7 points]** What is the probability that the tester shows a positive result for a (randomly selected) driver? (write your answer in terms of α, β_1, β_2).

     Solution: Let $P(\text{Positive})$ be the probability that the tester shows a positive result for a randomly selected driver. Using the law of total probability, we can express this as:

     $$P(\text{Positive}) = P(\text{Positive} \mid \text{High})P(\text{High}) + P(\text{Positive} \mid \text{Not High})P(\text{Not High})$$

     Given the problem statement:

     - $P(\text{High}) = \alpha$
     - $P(\text{Not High}) = 1 - \alpha$
     - $P(\text{Positive} \mid \text{High}) = 1 - \beta_1$
     - $P(\text{Positive} \mid \text{Not High}) = \beta_2$

     Thus, the probability of a positive result is:

     $$P(\text{Positive}) = \alpha(1 - \beta_1) + (1 - \alpha)\beta_2$$

- **[7 points]** The police have collected test results for n randomly selected drivers (i.i.d. samples). What is the likelihood that there are exactly n_+ positive samples among the n samples? Write your solution in terms of α, β_1, β_2, n_+ and n.

Solution: Each test result is an independent Bernoulli trial with success probability $P(\text{Positive})$. The likelihood function for observing n_+ positive results (out of n total tests) follows a binomial distribution:

$$P(n_+ \mid n, \alpha, \beta_1, \beta_2) = \binom{n}{n_+} \left(P(\text{Positive})\right)^{n_+} \left(1 - P(\text{Positive})\right)^{n-n_+}$$

Substituting $P(\text{Positive}) = \alpha(1-\beta_1) + (1-\alpha)\beta_2$, the likelihood is:

$$P(n_+ \mid n, \alpha, \beta_1, \beta_2) = \binom{n}{n_+} \left(\alpha(1-\beta_1) + (1-\alpha)\beta_2\right)^{n_+} \left(1 - (\alpha(1-\beta_1) + (1-\alpha)\beta_2)\right)^{n-n_+}$$

- **[10 points]** What is the maximum likelihood estimate of α given a set of n random samples from which n_+ are positive results? In this part, you can assume that β_1 and β_2 are fixed and given. Simplify your final result in terms of n, n_+, β_1, β_2.

Solution:

$$\alpha^{ML} = \arg\min_{\alpha} \Big[-\log P(n_+ \mid n, \alpha, \beta_1, \beta_2) \Big]$$

Since $\log \binom{n}{n_+}$ is independent of α, it can be omitted. We now differentiate the remaining log-likelihood with respect to α:

$$\frac{\partial}{\partial \alpha} \Big[n_+ \log\left(\alpha(1-\beta_1) + (1-\alpha)\beta_2\right) + (n-n_+) \log\left(1 - \alpha(1-\beta_1) - (1-\alpha)\beta_2\right) \Big] = 0$$

$$\implies n_+ \frac{1-\beta_1-\beta_2}{\alpha(1-\beta_1) + (1-\alpha)\beta_2} + (n-n_+) \frac{-(1-\beta_1)+\beta_2}{1-\alpha(1-\beta_1)-(1-\alpha)\beta_2} = 0$$

Now we assume $1-\beta_1-\beta_2 \neq 0$ (otherwise, the likelihood does not depend on the values of α, and the derivative with respect to α is always zero). Therefore, we can factor out $1-\beta_1-\beta_2$, and this simplifies to:

$$\frac{n_+}{\alpha(1-\beta_1) + (1-\alpha)\beta_2} = \frac{n-n_+}{1-\alpha(1-\beta_1)-(1-\alpha)\beta_2}$$

Cross-multiply:

$$n_+ \left(1 - \alpha(1-\beta_1) - (1-\alpha)\beta_2\right) = (n-n_+)\left(\alpha(1-\beta_1) + (1-\alpha)\beta_2\right)$$
$$\implies n_+ \left(1 - \alpha(1-\beta_1-\beta_2) - \beta_2\right) = (n-n_+)\left(\alpha(1-\beta_1-\beta_2) + \beta_2\right)$$

Move all terms involving α to one side:

$$n_+ - n_+\beta_2 - (n-n_+)\beta_2 = \alpha \Big[(n-n_+)(1-\beta_1-\beta_2) + n_+(1-\beta_1-\beta_2) \Big]$$

Simplify both sides:

$$n_+ - n\beta_2 = \alpha n (1-\beta_1-\beta_2)$$

Solve for α:

$$\alpha = \frac{n_+ - n\beta_2}{n(1 - \beta_1 - \beta_2)}$$

This is the maximum likelihood estimate (MLE) of α.

- **[6 points]** What will be the maximum likelihood estimate of α for the special cases of
  - (i) $\beta_1 = \beta_2 = 0$
  - (ii) $\beta_1 = \beta_2 = 0.5$
  - (iii) $\beta_1 = 0.2$, $\beta_2 = 0.3$

Solution:

(i) $\beta_1 = \beta_2 = 0$

When both $\beta_1 = \beta_2 = 0$, the test is perfectly accurate. The probability of a positive result is $P(\text{Positive}) = \alpha$. The MLE for α is simply:

$$\hat{\alpha} = \frac{n_+}{n}$$

(ii) $\beta_1 = \beta_2 = 0.5$

When both $\beta_1 = \beta_2 = 0.5$, the test is entirely random, as it gives a positive result half the time regardless of the driver's actual condition. In this case, the probability of a positive result is always 0.5, and the MLE for α is undefined since the likelihood does not depend on α.

(iii) $\beta_1 = 0.2$, $\beta_2 = 0.3$

Substituting these values into the MLE formula:

$$\hat{\alpha} = \frac{n_+ - 0.3n}{(1 - 0.2 - 0.3)n} = \frac{n_+ - 0.3n}{0.5n} = 2(n_+ - 0.3n)/n$$

Thus, the MLE for α in this case is:

$$\hat{\alpha} = 2(n_+ - 0.3n)/n$$