Homework 2 Part 1 - Solutions

This is an individual assignment.

Solve all problems by hand. You may type your answers in markdown cells or `push` a PDF file with your handwritten answers.

Problem 1 (15 points)

Consider a training set containing positive natural numbers - including zero - ($x \in \mathbb{N}_0$) for 2 classes, C_0 and C_1. The training set has 100 samples for class C_0 and 50 for C_1.

Suppose that you have reason to believe that samples belonging from C_0 are drawn from a Poisson random variable with parameter $\lambda > 0$, and samples belonging to C_1 are drawn from a Binomial random variable with parameters $n \in \mathbb{N}_0$ and $p \in [0,1]$. In other words:

$$p(x|C_0) = \frac{\lambda^x e^{-\lambda}}{x!}$$
$$p(x|C_1) = \binom{n}{x} p^x (1-p)^{n-x}$$

where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the binomial coefficient of "n choose x".

Consider $\lambda = 2$, $n = 10$ and $p = 0.5$. Consider the test point $x = 3$. Use the Bayesian classifier to assign $x = 3$ to C_0 or C_1? Show your work.

The prior probability for each class can be estimated as its relative frequency in the training set, namely

$$p(C_1) = \frac{100}{150} = \frac{2}{3} \text{ and } p(C_2) = \frac{50}{150} = \frac{1}{3}$$

For the provided parameter values, the data likelihood for class 1 and class 2 are defined as:

$$p(x|C_0) = \frac{2^3 e^{-2}}{3!} = \frac{4e^{-2}}{3}$$
$$p(x|C_1) = \binom{10}{3} 0.5^3 (1-0.5)^{10-2} = \frac{120}{2^{10}}$$

For the test point $x = 3$, we compute the posterior probability with

$$P(C_k|x=3) = \frac{p(x=3|C_k)p(C_k)}{p(x=3)} = \frac{p(x=3|C_k)p(C_k)}{\sum_{j=1}^{2} p(x=3|C_j)p(C_j)}$$

We now find:

$$P(C_0|x=3) = \frac{\frac{4e^{-2}}{3} \times \frac{2}{3}}{\frac{4e^{-2}}{3} \times \frac{2}{3} + \frac{120}{2^{10}} \times \frac{1}{3}} \approx 0.755$$

$$P(C_1|x=3) = \frac{\frac{120}{2^{10}} \times \frac{1}{3}}{\frac{4e^{-2}}{3} \times \frac{2}{3} + \frac{120}{2^{10}} \times \frac{1}{3}} \approx 0.245$$

Since $P(C_0|x=3) > P(C_1|x=3)$ then the test point $x=3$ is assigned to class C_0.

In [1]:

```
from scipy.special import binom

y0 = 4*np.exp(-2)/3
y1 = binom(10,3)/(2**10)

p0 = 2/3
p1 = 1-p0

pos0 = (y0*p0) / (y0*p0 + y1*p1)
pos1 = (y1*p1) / (y0*p0 + y1*p1)

pos0, pos1
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/var/folders/4p/y0h9xc8s5jq1s8b8591y3f5r0000gn/T/ipykernel_13457/663901916.py
in <module>
      1 from scipy.special import binom
      2 
----> 3 y0 = 4*np.exp(-2)/3
      4 y1 = binom(10,3)/(2**10)
      5 

NameError: name 'np' is not defined
```

Problem 2 (30 points)

Suppose you have a training set with N data points $\{x_i\}_{i=1}^{N}$, where $x_i \in \mathbb{Z}_0^+$ (set of nonnegative integers - $0, 1, 2, 3, \ldots$). Assume the samples are independent and identically distributed (i.i.d.), and each sample is drawn from a Geometric random variable with probability mass function:

$$P(x|\rho) = \rho(1-\rho)^x$$

Moreover, consider the Beta density function as the prior probability on the success probability, ρ,

$$P(\rho|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\rho^{\alpha-1}(1-\rho)^{\beta-1}$$

Answer the following questions:

1. (8 points) **Derive the maximum likelihood estimate (MLE) for the rate parameter ρ. Show your work.**
2. (10 points) **Derive the maximum a posteriori (MAP) estimate for the rate parameter ρ. show your work.**
3. (6 points) **Is the Beta distribution a conjugate prior for the success probability, ρ, of the Geometric distribution? Why or why not?**
4. (6 points) **Suppose you would like to update the Beta prior distribution and the MAP point estimation in an online fashion, as you obtain more data. Write the pseudo-code for the online update of the prior parameters. In your answer, specify the new values for the parameters of the prior.**

1. The observed data likelihood is:

$$\mathcal{L}^0 = \prod_{i=1}^{N} \rho(1-\rho)^{x_i} = \rho^N (1-\rho)^{\sum_{i=1}^N x_i}$$

The log-likelihood is given by:

$$\mathcal{L} = \ln \mathcal{L}^0 = N\ln(\rho) + \left(\sum_{i=1}^{N} x_i\right)\ln(1-\rho)$$

We can now find the MLE estimation for λ:

$$\frac{\partial \mathcal{L}}{\partial \rho} = 0 \iff \frac{N}{\rho} - \frac{\sum_{i=1}^N x_i}{1-\rho} = 0 \iff N - N\rho - \rho\sum_{i=1}^{N} x_i \iff \rho_{MLE} = \frac{N}{N + \sum_{i=1}^N}$$

1. The observed data likelihood for MAP is:

$$\begin{aligned}\mathcal{L}^0 &\propto \left(\prod_{i=1}^{N} \rho(1-\rho)^{x_i}\right)\rho^{\alpha-1}(1-\rho)^{\beta-1} \\ &= \rho^N(1-\rho)^{\sum_{i=1}^N x_i}\rho^{\alpha-1}(1-\rho)^{\beta-1} \\ &= \rho^{N+\alpha-1}(1-\rho)^{\sum_{i=1}^N x_i + \beta - 1}\end{aligned}$$

The log-likelihood is given by:

$$\mathcal{L} = (N + \alpha - 1)\ln(\rho) + \left(\sum_{i=1}^{N} x_i + \beta - 1\right)\ln(1-\rho)$$

We can now find the MAP estimation for λ:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \rho} = 0 \iff & \frac{N+\alpha-1}{\rho} - \frac{\sum_{i=1}^{N} x_i + \beta - 1}{1-\rho} = 0 \\
\iff & N + \alpha - 1 - \rho\left(N + \alpha + \sum_{i=1}^{N} x_i + \beta - 2\right) = 0 \\
\iff & \rho_{MAP} = \frac{N+\alpha-1}{\sum_{i=1}^{N} x_i + N + \alpha + \beta - 2}
\end{aligned}$$

1. From the previous problem, we compute the posterior probability as:

$$\mathcal{L}^0 = \rho^{N+\alpha-1}(1-\rho)^{\sum_{i=1}^{N} x_i + \beta - 1}$$

We see that this posterior is proportionally equal to the parametric form of the prior distribution $P(\rho|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\rho^{\alpha-1}(1-\rho)^{\beta-1}$, where

$$\begin{aligned}
\alpha &\longleftarrow \alpha + N \\
\beta &\longleftarrow \beta + \sum_{i=1}^{N} x_i
\end{aligned}$$

1. The pseudo-code for online update of the prior is as follows:
   A. Start at iteration $t = 0$. Initialize the prior parameters $\alpha^{(t)}$ and $\beta^{(t)}$.
   B. Compute the parameter estimation for the current prior probability:

$$\lambda_{\text{MAP}} = \frac{N+\alpha-1}{\sum_{i=1}^{N} x_i + N + \alpha + \beta - 2}$$

   C. Update the prior parameters

$$\begin{aligned}
\alpha^{(t+1)} &\leftarrow \alpha^{(t)} + N \\
\beta^{(t+1)} &\leftarrow \beta^{(t)} + \sum_{i=1}^{N} x_i
\end{aligned}$$

   D. Increment iteration counter

$$t \leftarrow t + 1$$

On-Time (5 points)

Submit your assignment before the deadline.

Submit Your Solution

Confirm that you've successfully completed the assignment.

Along with the Notebook, include a PDF of the notebook with your solutions.

`add` and `commit` the final version of your work, and `push` your code to your GitHub repository.

Submit the URL of your GitHub Repository as your assignment submission on Canvas.
