Homework 2 - Machine Learning

Professor: Daniel S. Menasché

1 Parameter Estimation: Likelihood, MAP and Fully Bayesian

The following list of 20 points

```
x=[1,3,2,2,3,0,1,2,1,2,1,1,2,2,1,3,3,2,1,3]
```

was generated by sampling a binomial distribution, with parameters N and 1/5,

```
#  R=binornd(N,1/5,1,20)
```

The goal of this homework is to estimate N.

1.1 Warming up

Using Matlab or any other program to work with matrices,

- plot a bar graph where the x axis are the values in the list and the y axis is the number of occurrences of each value
- plot the empirical probability mass function (obtained from the item above after normalization)
- plot the probability mass function of a binomial distribution with parameters
  - N=5, p=1/5
  - N=10, p=1/5

- N=20, p=1/5

By visual inspection, which of the above you think is more likely to have generated the list of 20 points?

1.2 Maximum Log Likelihood

- plot the likelihood function of the 20 points, as a function of the parameter N. Your graph should have as the x axis integer numbers from 1 to 20, and the y axis $p(x|N)$ (recall that x is the list of 20 points).

  Hint: the Matlab function *binopdf* might be useful.

- plot the log of the likelihood, i.e., the log of the function you ploted above

What is the value of N that maximizes the likelihood? What about the log likelihood?

1.3 Maximum Posterior

Suppose we are given a *prior* distribution for N. We are told that N is equal to 10 with probability 0.1, and is equal to $1, 2, \ldots, 9, 11, 12, \ldots, 20$ with probability 0.9/19.

- plot the likelihood times the prior for each of the values of N. Is the function plotted a probability mass function?

- plot the posterior (you can obtain the posterior by normalizing the curve above). Is the function plotted a probability mass function?

What is the value of N that maximizes each of the 2 curves obtained above? Compare this value of N with the one you obtained before normalizing.

1.4 Fully Bayesian

Assume that we got additional 10 samples of the distribution,

```
4    2    0    2    4    0    1    1    1    2
```

true distr. / classified as from...	distribution 1	distribution 2
distribution 1	0	10
distribution 2	1000	0

Tabela 1: Loss function

How can we use these extra samples to refine our estimate of N?

Let us denote the posterior obtained in the previous item by P. Using P as the new prior, recompute the new posterior using the samples above.

Describe a general method that allows you to refine your estimate of N accounting for new samples as they arrive.

2 Classification

Consider now two distributions

- **distribution 1**: the distribution that generated the 20 samples in part 1
- **distribution 2**: a Binomial distribution with parameters $N = 20$ and $p = 1/5$

Answer the following questions:

- we are given a sample equal to 2. Is it more likely to have been generated by distribution 1 or 2? Use maximum likelihood estimation to support your answer.
- we are given a sample equal to 2 and the loss table shown in Table 1. Is it more advisable to classify this sample as coming from distribution 1 or from distribution 2? Justify.
- we are given a sample equal to 2 and the loss table shown in Table 1. Plus, we have the following prior information: a reliable source mentioned that samples come with probability 0.9 from distribution 1 and with probability 0.1 from distribution 2. In light of this prior information, is it more advisable to classify the point as coming from distribution 1 or 2? Justify.