Machine Learning Worksheet 4

Maximum Likelihood Estimation

1 Still refreshing ...

Problem 1. An unbiased coin is flipped until one head is thrown. What is the expected number of tails and the expected number of heads? Show your work.

Problem 2. There are eleven urns labeled by $u \in \{0, 1, 2, \ldots, 10\}$, each containing ten balls. Urn u contains u black balls and $10 - u$ white balls. Alice selects an urn u at random and draws N times with replacement from that urn, obtaining n_B black balls and $N - n_B$ white balls. If after $N = 10$ draws $n_B = 3$ black balls have been drawn, what is the probability that the urn Alice is using is urn u?

Now, let Alice draw another ball from the same urn. What is the probability that the next drawn ball is black (show your work)?

2 Parameter Estimation

Consider n samples $x_1, \ldots, x_n$ drawn independently and identically (i.i.d.) from a given distribution $P(X|\theta)$. This distribution is usually parametrized (e.g. one parameter representing its mean, one its variance, etc.); these parameters are denoted by θ. One wants to find accurate estimates for these parameters using the n samples only. *Maximum Likelihood Estimation* (MLE) finds estimates for the various parameters at hand by maximizing the likelihood $P(x_1, x_2, \ldots, x_n|\theta) = \Pi_{i=1}^{n} P(x_i|\theta)$. (i.e. the probability of observing the n samples at hand). Note that usually one considers the *log likelihood*, $\log P(x_1, \ldots, x_n|\theta)$.

2.1 Coins

Let X be a Bernoulli random variable. The Bernoulli distribution is only parametrized by one parameter, $\theta = P(X = 1)$.

Problem 3. For n i.i.d. observations of X determine the MLE for θ. You might want to use $P(X = x|\theta) = \theta^x(1 - \theta)^{1-x}$.

Now we look at slightly more complex distribution, the binomial distribution.

Problem 4. Consider a Bernoulli random variable X and suppose we have observed m occurences of $X = 1$ and l occurences of $X = 0$ in a sequence of $m + l$ Bernoulli experiments. We are only interested in the number of occurences of $X = 1$ – we will model this with a Binomial distribution with parameter μ. A prior distribution for μ is given by the Beta distribution. Show that the posterior *mean* value of

μ lies between the prior mean of μ and the maximum likelihood estimate for μ. To do this, show that the posterior mean can be written as λ times the prior mean plus $(1-\lambda)$ times the maximum likelihood estimate, with $0 \leq \lambda \leq 1$. This illustrates the concept of the posterior mean being a compromise between the prior distribution and the maximum likelihood solution.

Note: The probability mass function for the Binomial distribution is defined as follows:

$$p(x = m|N, \mu) = \binom{N}{m} \mu^m (1-\mu)^{N-m}$$

2.2 Poisson distribution

Let X be Poisson distributed.

Problem 5. Again, for n i.i.d. samples from X, determine the maximum likelihood estimate for λ. Show that this estimate is unbiased!

In class we also talked about avoiding overfitting of parameters via *prior* information. Compute the posterior distribution over λ, assuming a $Gamma(\alpha, \beta)$ prior for it. Compute the MAP for λ under this prior. Show your work.