Machine Learning Worksheet Solution 3

Julius Jankowski

1 Optimizing Likelihoods: Monotonic Transforms

Problem 1:

Derivations for $p(\mathbf{x} \mid \theta)$:

$$p(\mathbf{x} \mid \theta) = \theta^t (1-\theta)^h \tag{1.1}$$

$$\frac{\partial p(\mathbf{x} \mid \theta)}{\partial \theta} = \theta^{t-1}(1-\theta)^{h-1}[t(1-\theta) - h\theta] \tag{1.2}$$

$$\frac{\partial^2 p(\mathbf{x} \mid \theta)}{\partial \theta^2} = \theta^{t-2}(1-\theta)^{h-2}[(t-1)t(1-\theta)^2 + (h-1)h\theta^2 - 2th\theta(1-\theta)] \tag{1.3}$$

Derivations for $log_b p(\mathbf{x} \mid \theta)$:

$$log_b p(\mathbf{x} \mid \theta) = t \cdot log_b(\theta) + h \cdot log_b(1-\theta) \tag{1.4}$$

$$\frac{\partial log_b[p(\mathbf{x} \mid \theta)]}{\partial \theta} = \frac{1}{ln(b)}[\frac{t}{\theta} - \frac{h}{1-\theta}] \tag{1.5}$$

$$\frac{\partial^2 log_b[p(\mathbf{x} \mid \theta)]}{\partial \theta^2} = -\frac{1}{ln(b)}[\frac{t}{\theta^2} + \frac{h}{(1-\theta)^2}] \tag{1.6}$$

Problem 2:

According to the chain derivation rule, the partial derivative of $ln(f(\theta))$ w.r.t. θ can be expressed like this:

$$\frac{\partial ln(f(\theta))}{\partial \theta} = \frac{\partial f(\theta)}{\partial \theta} \cdot \frac{1}{f(\theta)} \tag{1.7}$$

The first condition for θ to maximize the function is to satisfy the following equation:

$$\frac{\partial f(\theta)}{\partial \theta} \cdot \frac{1}{f(\theta)} = 0 \tag{1.8}$$

This implies that either $\frac{\partial f(\theta)}{\partial \theta}$ or $\frac{1}{f(\theta)}$ must be zero, which is equal to the formulation when looking for the maximum of $\frac{\partial f(\theta)}{\partial \theta}$ since $\frac{1}{f(\theta)}$ will not become zero for any real θ.

Since the search for the maximum can be simplyfied by taking the logarithm of the probability, it is recommended to use this for maximum likelihood search.

2 Properties of MLE and MAP

Problem 3:

When we are searching for the MAP solution for θ, we have to optimize the conditional probability:

$$p(\theta \mid D) = p(D \mid \theta)\frac{p(\theta)}{p(D)} \tag{2.1}$$

Since we are looking for the root in the conditional probability, every $p(\theta)$ which is constant w.r.t. θ does not change the maximum solution. And since $p(D)$ does not depend on θ as well, the solution for ML and MAP is equal in this particular case.

Problem 4:

Posterior distribution is computed as following:

$$p(\theta \mid D) = p(x = m \mid N, \theta)\frac{p(\theta)}{p(D)} \tag{2.2}$$

Inserting the binomial and beta distribution respectively:

$$p(\theta \mid D) = c \cdot \theta^m (1-\theta)^{N-m} \cdot \theta^{a-1}(1-\theta)^{b-1} = c \cdot \theta^{m+a-1}(1-\theta)^{N-m+b-1} \tag{2.3}$$

with a factor c which is constant w.r.t. θ. This corresponds to a beta distribution in θ:

$$p(\theta \mid D) \propto Beta(p = m + a, q = N - m + b) \tag{2.4}$$

The according expectation value is:

$$E[\theta \mid D] = \frac{m+a}{N+a+b} = \frac{m}{N+a+b} + \frac{a}{N+a+b} \tag{2.5}$$

which can be splitted into a term from the expectation $E[\theta]$ based on the prior distribution and a term from the maximum likelihood solution θ_{ML}:

$$E[\theta \mid D] = \lambda \cdot E[\theta] + (1-\lambda) \cdot \theta_{ML} = \lambda \frac{a}{a+b} + (1-\lambda)\frac{m}{N} \tag{2.6}$$

with $\lambda = \frac{a+b}{N+a+b}$.

3 Poisson Distribution

Problem 5:

Under the assumption of having n i.i.d. samples of a Poisson distributed variable X, we have a likelihood to observe that data given the Poisson parameter λ:

$$p(\mathbf{x} \mid \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \lambda^{|\mathbf{x}|_1} \cdot e^{-n\lambda} \prod_{i=1}^{n} \frac{1}{x_i!} \tag{3.1}$$

Computing the maximum likelihood solution for λ based on the log-likelihood:

$$\frac{\partial ln[p(\mathbf{x} \mid \lambda)]}{\partial \lambda} = \frac{1}{\lambda} \cdot \sum_{i=1}^{n} x_i - n = 0 \tag{3.2}$$

$$\lambda_{ML} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.3}$$

which is basically the arithmetic mean of the n samples.

Applying all known distributions to the posterior distribution of λ:

$$p(\lambda \mid \mathbf{x}) = p(\mathbf{x} \mid \lambda) \cdot \frac{p(\lambda)}{p(\mathbf{x})} = c \cdot \lambda^{|\mathbf{x}|_1 + p - 1} \cdot e^{-\lambda(n+b)} \tag{3.4}$$

Here we can use the result for the general ML solution for Poisson distributed samples because the distribution can be interpreted as if the number of samples have been extended by b and the resulting sum over all samples have been extended by $p - 1$. Hence the resulting MAP solution is:

$$\lambda_{MAP} = \frac{p - 1 + \sum_{i=1}^{n} x_i}{n + b} \tag{3.5}$$