

UNC BIOS 735

SUMMARY

## Creating my own project

*Author:*  
Mingwei Fei

*Special colaborator:*  
Bonnie

Jan, 2023



# Contents



# Chapter 1

## EM Algorithm

### 1.1 EM Algorithm

#### 1.1.1 Jessen's Inequality

In a convex function, we have  $E[f(x)] \geq f[E(X)]$ . Associate this with the graph, the average of two points is high above the point in the middle.

This could be related to the likelihood function, as the optimal point is between two points, that we can approximate the likelihood of the optimal point by average the likelihood of the two points.

But log-likelihood function is concave, while not convex.

#### 1.1.2 Application in EM Example

Consider an experiment with coin A that has a probability  $\theta_A$  of heads, and a coin B that has a probability  $\theta_B$  of tails. We draw  $m$  samples as follows - for each sample, pick one of the coins at random, flip it  $n$  times, and record the number of heads and tails (that sum to  $n$ ). If we recorded which coin we used for each sample, we have complete information and can estimate  $\theta_A$  and  $\theta_B$  in closed form. To be very explicit, suppose we drew 5 samples with the number of heads and tails represented as a vector  $x$ , and the sequence of coins chosen was A,A,B,A,B.

- (i) Complete Information with Maximum Likelihood Then the complete log likelihood is

$$\log p(X, \theta) = \log p(x_1, \theta_A) + \log p(x_2, \theta_A) + \log p(x_3, \theta_B) + \log p(x_4, \theta_A) + \log p(x_5, \theta_B)$$

We will use  $z_i$  to indicate the label of the  $i$ th coin, that is - whether we used coin A or B to generate the  $i$ th sample.

- (ii) Incomplete information However, if we did not record the coin we used, we have missing data and the problem of estimating  $\theta_B$  is harder to solve. One way to approach the problem is to ask - can we assign weights  $w_i$  to each sample according to how likely it is to be generated from coin A or coin B?

Remark: It is a common practice dealing with missing discrete variables, which use the posterior probability based on the data. That is how the Jensen's inequality come in the play. The missing information is an expectation in the log-likelihood function, so the expectation of the log-likelihood function is smaller than the log-likelihood function of the expected missing information.

With knowledge of  $w_i$ , we can maximize the likelihood to find  $\theta$ . Similarly, given  $w_i$ , we can calculate what  $\theta$  should be. So the basic idea behind Expectation Maximization (EM) is simply to start with a guess for  $\theta$ , then calculate z, then update  $\theta$  using this new value for z, and repeat till convergence. The derivation below shows why the EM algorithm using this "alternating" updates actually works.

A verbal outline of the derivation - first consider the log likelihood function as a curve (surface) where the base is  $\theta$ . Find another function Q of  $\theta$  that is a lower bound of the log-likelihood but touches the log likelihood function at some  $\theta$  (E-step). Next find the value of  $\theta$  that maximizes this function (M-step). Now find yet another function of  $\theta$  that is a lower bound of the log-likelihood but touches the log likelihood function at this new  $\theta$ . Now repeat until convergence - at this point, the maxima of the lower bound and likelihood functions are the same and we have found the maximum log likelihood. See illustration below.

- (iii) E- Step

In the E-step, we identify a function which is a lower bound for the log-likelihood. How do we choose the distribution  $Q_i$ ? We want the Q function to touch the log-likelihood, and know that Jensen's inequality is an equality only if the function is constant. So  $Q_i$  is just the posterior distribution of  $z_i$ , and this completes the E-step.

In any case,

$$[\phi^N - \phi][f_1(x) - kf_0(x)] \geq 0$$

and therefore,

$$\begin{aligned} \int [\phi^N - \phi][f_1(x) - kf_0(x)]dv &\geq 0 \\ \int [\phi^N - \phi]f_1(x)dv &\geq \int [\phi^N - \phi]kf_0(x)dv \end{aligned}$$

The left-hand side is  $E_1(\phi^N) - E_1(\phi)$  and the right-hand side

$$k(E_0[\phi^N] - E_0[\phi]) = k(\alpha - E_0[\phi]) \geq 0$$

## Chapter 2

# MCMC

### 2.1 Rejection Sampling

The underlying principle is to keep the sampling points based on the probability ratio between the targeted probability and the known probability (known prob is larger than the targeted distribution). For example, if the ratio is high, then we have high probability keeping the sampling points. Otherwise, we will reject the sampling points a lot.

The Monte Carlo shows that the distribution will be stationary when there is large enough sampling points. Although the way we select the sampling points based on the previous sampling point, eventually we will get to the stationary distribution.

#### 2.1.1 Mathematical Function

$f(Y)$  is the targeted distribution, and  $g(Y)$  is the known larger distribution.  $U \sim U(0, 1)$ , and the value of  $y$  each time is generated under the density function  $g(\cdot)$  of the proposal distribution  $Y$ .

$$\begin{aligned}
p(U \leq \frac{f(Y)}{Mg(Y)}) &= E\left[I\left(U \leq \frac{f(Y)}{Mg(Y)}\right)\right] \\
&= E\left[EI\left(U \leq \frac{f(Y)}{Mg(Y)}\right)|Y\right] \\
&= E\left[P\left(U \leq \frac{f(Y)}{Mg(Y)}\right)|Y\right] \\
&= E\left[\frac{f(Y)}{Mg(Y)}\right] \quad P(U \leq u) = u \\
&= \int_{y:g(y)>0} \frac{f(Y)}{Mg(Y)} g(y) dy \\
&= \frac{1}{M} \int_{y:g(y)>0} f(y) dy \\
&= \frac{1}{M}
\end{aligned}$$



# Bibliography

- [1] Armstrong, M.A. *Basic Topology*. England: Editorial Board, 2000.
- [2] Coxeter, H.S.M. *Introduction to Geometry*. Toronto: John Wiley and Sons Inc, 1969.
- [3] Chronister, James. *Blender Basics*. Estados Unidos: 2004.
- [4] Engelking, Ryszard. *General Topology*. Berlin: Hildermann, 1989.
- [5] Engelking, Ryszard y Karol Sieklucki. *Topology: A Geometric Approach*. Berlin: Hildermann, 1992.
- [6] Hatcher, Allen. *Algebraic Topology*. Cambridge: Cambridge University Press, 2002.
- [7] Hoffman, Kenneth y Ray Kunze. *Linear Algebra, second edition*. New Jersey: Prentice Hall. Inc, 1961.
- [8] Kackzynski, T., K. Mischaikow y M. Mrozen. *Algebraic Topology: A Computational Approach*. 2000.
- [9] Kosniowski, Czes. *Topología Algebraica*. Barcelona: Editorial Revert, 1986.
- [10] Lefschetz, Solomon. *Algebraic Topology*. Rhode Island: American Mathematical Society, 1991.
- [11] May, J. P. *A Concise Course in Algebraic Topology*. Chicago.
- [12] Mill, J. Van. *Infinite-Dimensional Topology*. Amsterdam: Elsevier Science Publications B. V, 1989.
- [13] Munkres, James R. *Topology*. New York: Prentice Hall Inc, 2000.
- [14] Roseman, Dennis. *Elementary Topology*. New Jersey: Prentice Hall Inc, 1999.
- [15] Sato, Hajime. *Algebraic Topology: An Intuitive Approach*. Rhode Island: American Mathematical Society, 1996.
- [16] Zomorodian, Afra J. *Topology for Computing*. United States: Cambridge University Press, 2005.