Expectation Maximum Algorithm

# 1. Maximum Likelihood Estimation

## Motivation

* We have data points drawn from set X
* We have parameter (parameter space)
* We have distribution so that
* Assume that we have drawn **independently identically** from distribution for
* The likelihood will be
* So with **“given”** dataset we want to maximum likelihood function so that the distribution is “close” to your “given” dataset

## Example

* Let start with simple example that we flip the coin.
* so data points is sequence of heads or tails. In this example **HHTTHHHTHH**
* is single parameter that probability of coin coming up heads.
* will be defined as
* n is number of your data points
* The problem that we want to find that can fit with your example? We try to maximize

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
| Derivative to find maximum point | => |  |
|  | <=> |  |

# 2. Expectation Maximum

## Motivation

* The problem is if your distribution is mixture of K Gaussian
* How can we use apply maximum likelihood in Gaussian Mixture Model with numbers parameter can be over 100 parameters? How’s about in case our model is formulated in term of “observed” and “unobserved” data. “Unobserved” in this case refer to **quantities.** For example, in given data points you don’t know gaussian model that your sample is drawn from. If we can measure them, we can estimate the parameters by maximum likelihood. That’s why Expectation Maximum is used to solve this case

## Algorithm

* From Gaussian mixture model, we will have likelihood function like:
* If is known, we can estimate . That means you have data points for specific distribution
* If is known, we can estimate . For example if , so that means is likely drawn from rather than
* That intuitive idea of Expectation Maximum. We **iterate:**
  + Expectation step: we calculate expect value with given parameters
  + Maximum step: calculate “MLE” of parameter given

## Mathematical Understanding

* If is known, unknown. Your MLE function will be
* If we also know , consider
* Now, we don’t know . We maximum *expected likelihood* of visible data where expectation is over distribution of hidden data

### E-step: Find

* Assume known (from previous iteration or initial )
* are event that drawn from
* D is observed datum
* Expect value of for each

### M-step: Reestimate

# 3. Some example for EM

## Three coin problem

* We have 3 coins. The problem is given sequence of Head and Tail from tossing coin 1,2 under condition. If coin 0 tossed before is H, we will toss coin 1 and If coin 0 is T, we toss coin 2.
* Define the problem
* Our partially observed data [H,T,H,T,H] with initial parameter

### E-step:

* With defined parameter
* After filling hidden variables for each sample {H,T,H,T,H} we will have

|  |  |
| --- | --- |
| 3 <H> |  |
|  |
| 2 <T> |  |
|  |

### M-step:

* New estimate for parameter:
* We continue until it converge