Mixture models - exercises

Exercise 1. Social group reasoning

Our knowledge about the social world is *structured*: we do not just know a collection of facts about particular people but believe that there are *kinds* of people with shared properties. Even infants make strong assumptions about people based on what language they speak, what foods they eat, and what actions they take (check out Katherine Kinzler's work!) How do we learn this structure at the same time as we are learning about individual people? In this exercise you will explore mixture models as a formal model of how we reason about social groups.

a)

Imagine you go to an alien planet and see 10 aliens: you notice three clear properties, some have antennae, some are green, and some make a distinctive 'blargh' noise. Implement a simple model assuming that there are two kinds of aliens with different distributions over these properties, but you have a priori uncertainty over what the distributions are, and whether any particular alien belongs to group A or group B.

HINT: each data point we observed in the chapter only had one property (from k different values). Here each alien has three properties. This means we need a way of observing all three properties under their respective prototype priors.

```
x
```

```
///fold:
```

```
...
```

```
var properties = ['antennae', 'green', 'blarghNoise']
```

```
var data = [
```

```
  {antennae : false, green: false, blarghNoise: false},
```

```
  {antennae : true,  green: true,  blarghNoise: true},
```

```
  {antennae : true,  green: true,  blarghNoise: true},
```

```
  {antennae : true,  green: true,  blarghNoise: true},
```

```
  {antennae : false, green: false, blarghNoise: false},
  {antennae : true,  green: true,  blarghNoise: true},
  {antennae : false, green: false, blarghNoise: false},
  {antennae : true,  green: true,  blarghNoise: true},
  {antennae : false, green: false, blarghNoise: false},
  {antennae : false, green: false, blarghNoise: false}
]

// Todo: sampleGroupPrototype takes a group and returns an object
// with property / probability pairs. E.g. {antannae: 0.2, green: 0.3, blarghNoise: 0.9}
// *Hint* lodash _.zipObject is useful for building dictionaries!
var sampleGroupPrototype = mem(function(groupName) {
  // Your code here...
})

var results = Infer({method: 'MCMC', kernel: {HMC: {steps: 10, stepSize: .01}},
                     samples: 3000}, function(){
  mapData({data: data}, function(datum) {

    // Your code here...
```

```
    mapData({data: properties}, function(property) {

      observe( // Your code here... )

    })

  })

  return {group1: sampleGroupPrototype('group1'),

          group2: sampleGroupPrototype('group2')}

})

viz.bar(properties, map(expectationOver(results, 'group1'), properties))

viz.bar(properties, map(expectationOver(results, 'group2'), properties))
```

run ▼

b)

Now imagine you hear a noise from inside a crater but you cannot see the alien that emitted it; this is a noisy observation. How can you use the model you learned above to make an educated guess about their other features?

Exercise 2: Detecting cheating

This problem is adapted from Section 6.5 of Lee \& Wagenmakers (2013) (https://faculty.washington.edu/jmiyamot/p548/leemd%20bayesian%20cog%20modeling%20-%20practical%20crs.pdf).

Consider the practical challenge of detecting if people cheat on a test. For example, people who have been in a car accident may seek financial compensation from insurance companies by feigning cognitive impairment such as pronounced memory loss. When these people are confronted with a memory test that is intended to measure the extent of their impairment, they may deliberately under-perform. This behavior is called malingering, and it may be accompanied by performance much worse than that displayed by real amnesiacs. Sometimes, for example, malingerers may perform substantially below chance.

Malingering is not always easy to detect, but is naturally addressed by a mixture model. Using this approach, it is possible to infer which of two categories – those who malinger, and those who are truthful or bona fide – each person belongs to, and quantify the confidence in each of these classifications. We consider an experimental study on malingering, in which each of p = 22 participants completed a memory test (Ortega, Wagenmakers, Lee, Markowitsch, & Piefke, 2012). One group of participants was told to

do their best. These are the bona fide participants. The other group of participants was told to under-perform by deliberately simulating amnesia. These are the malingerers. Out of a total of n = 45 test items, the participants get 45, 45, 44, 45, 44, 45, 45, 45, 45, 45, 30, 20, 6, 44, 44, 27, 25, 17, 14, 27, 35, and 30 correct. Because this was an experimental study, we know that the first 10 participants were bona fide and the next 12 were instructed to malinger.

a)

Implement a simple mixture model inferring which group each participant belongs to. Examine the posteriors over group-level parameters.

HINT: the group-level variables you are trying to infer are the error rates; it probably makes sense to assume that the malingerers are worse than bonafide participants, but have uncertainty over each value.

```
var scores = [45, 45, 44, 45, 44, 45, 45, 45, 45, 45, 30, 20, 6, 44, 44, 27, 25, 17, 14, 27, 3

var subjIDs = _.range(scores.length)

var data = map(function(datum) {return _.zipObject(['subjID', 'score'], datum)}, _.zip(subjIDs

var inferOpts = {method: 'MCMC', samples: 10000}

var results = Infer(inferOpts, function() {

  // Your code here...

  var obsFn = function(datum){

    observe(// Your code here...)

  }

  mapData({data: data}, obsFn)

  // Your code here...
```

```
  return // Your code here...
})

viz.marginals(results)
```

run ▼

b)

Examine the posteriors over group membership for each participant. Did all of the participants follow the instructions? (i.e. are the first 10 inferred to be in one group and the next 12 in the other?)