Two component Gaussian models
Created by: AJ Richards
Last updated: July 16, 2014

Contents

0.1 Problem definition

In the book Elements of Statistical Learning they define a two component Gaussian problem. Here we use the same data.

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.mlab as mlab
import pymc as mc

book_mu1 = 4.62
book_mu2 = 1.06
book_var1 = 0.87
book_var2 = 0.77
book_pi = 0.546

x1 = np.array([-0.39,0.12,0.94,1.67,1.76,2.44,3.72,4.28,4.92,5.53])
x2 = np.array([ 0.06,0.48,1.01,1.68,1.80,3.25,4.12,4.60,5.28,6.22])
x = np.hstack([x1,x2])
```

```
...
```

0.2 Expectation maximization algorithm

```
from TwoComponentGaussEM import TwoComponentGaussianEM

numIters = 25
numRuns = 20
tcg = TwoComponentGaussianEM(x, numIters, numRuns,verbose=False)

print(’maxLike␣%s’%tcg.maxLike)
print(’bestEstimates␣%s’%(tcg.bestEst))
```

```
...
maxLike -40.5132732737
bestEstimates {’sig1’: 0.80186156702591549, ’sig2’: 0.82962570734635754, ’mu1’: 1.0771558178573672, ’mu2’:
     4.648719485769476, ’pi’: 0.44723945326625514, ’n’: 20}
```

```
fig = plt.figure(1)
ax = fig.add_subplot(111)
n,bins,patches = ax.hist(x,15,normed=1,facecolor=’gray’,alpha=0.75)

p1 = mlab.normpdf(bins, book_mu1, np.sqrt(book_var1))
p2 = mlab.normpdf(bins, book_mu2, np.sqrt(book_var2))
l1 = plt.plot(bins, p1, ’r--’, linewidth=1)
l2 = plt.plot(bins, p2, ’r--’, linewidth=1)

## add a ’best fit’ for EM
p3 = mlab.normpdf( bins, tcg.bestEst[’mu1’], np.sqrt(tcg.bestEst[’sig1’]))
p4 = mlab.normpdf( bins, tcg.bestEst[’mu2’], np.sqrt(tcg.bestEst[’sig2’]))
l3 = plt.plot(bins, p3, ’k-’, linewidth=1)
l4 = plt.plot(bins, p4, ’k-’, linewidth=1)

#p3 = mlab.normpdf(bins, est_mu1, np.sqrt(est_var1))
#p4 = mlab.normpdf(bins, est_mu2, np.sqrt(est_var2))
#l3 = plt.plot(bins, p3, ’k-’, linewidth=1)
#l4 = plt.plot(bins, p4, ’k-’, linewidth=1)


## add a ’best fit’ for pymc
#p3 = mlab.normpdf(bins, est_mu1, np.sqrt(est_var1))
#p4 = mlab.normpdf(bins, est_mu2, np.sqrt(est_var2))
#l3 = plt.plot(bins, p3, ’k-’, linewidth=1)
#l4 = plt.plot(bins, p4, ’k-’, linewidth=1)

plt.xlabel(’x’)
plt.ylabel(’freq’)
plt.ylim([0,0.8])

ax.legend((l1[0], l3[0]), (’Book␣Estimate’, ’My␣Estimate’) )
plt.savefig(’TwoComponentGaussSummary.png’)
```

```
...
```

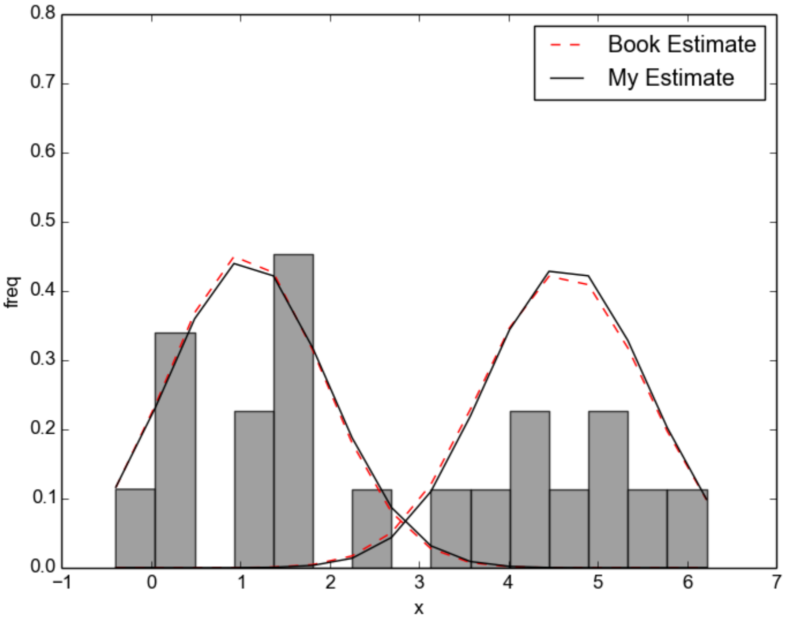

Figure 1: Model summary.