Scientific Computing for Biologists

Lecture 10: Mixture Models and Multi-dimensional Scaling

Instructor: Paul M. Magwene

05 November 2013

Outline of Lecture

- K-means clustering
- Mixture model based clustering
- Multi-dimensional scaling (MDS)

Gaussian Mixture Models

A common starting point in mixture modeling is to assume that the components are Gaussian.

If the data are univariate, then the mixture model is given by:

$$p_{\text{mix}} = \sum_{s=1}^{g} \pi_s f(\boldsymbol{x}|\mu_i, \sigma_i^2)$$

where the μ_i and σ_i are the means and standard deviations of each component distribution and:

$$f(\boldsymbol{x}|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

MDS Example: Road Distances between U.S. Cities

	BOS	CHI	DC	DEN	LA	MIA	NY	SEA	SF
BOS	0	963	429	1949	2979	1504	206	2976	3095
CHI	963	0	671	996	2054	1329	802	2013	2142
DC	429	671	0	1616	2631	1075	233	2684	2799
DEN	1949	996	1616	0	1059	2037	1771	1307	1235
LA	2979	2054	2631	1059	0	2687	2786	1131	379
MIA	1504	1329	1075	2037	2687	0	1308	3273	3053
NY	206	802	233	1771	2786	1308	0	2815	2934
SEA	2976	2013	2684	1307	1131	3273	2815	0	808
SF	3095	2142	2799	1235	379	3053	2934	808	0

Some Code

Here's some code:

```
def myfunc(x):
    print "Hello, Python World"

def yourfunc(x):
    print "Hello, LaTeX World!"

x = 10
myfunc(x)
```

Last Slide

This is the *last* slide of the document.

It was created using:

- YASnippet
- AucTeX
- XeLaTeX