Running sLDA

We are running an implementation of sLDA from Wang, which predicts categorical labels (consistent with our tasks so far).

http://www.cs.cmu.edu/~chongw/slda/

Another implementation is here but it handles continuous valued labels only?

http://cran.r-project.org/web/packages/lda/lda.pdf

General LDA info:

http://www.cs.princeton.edu/~blei/topicmodeling.html

The executable (after compiling according to instructions).

In [1]:
```
!slda
```

```
usage: slda [est] [data] [label] [settings] [alpha] [k] [random/seeded/model_path]
[directory]
       slda [inf] [data] [label] [settings] [model] [directory]
```

Input data

From the README:

[data] is a file where each line is of the form:

```
[M] [term_1]:[count] [term_2]:[count] ...  [term_N]:[count]
```

where [M] is the number of unique terms in the document, and the [count] associated with each term is how many times that term appeared in the document.

A provided sample training data file.

In [7]:
```
!head -n 2 Wang/sample_data/Labelme/images/train-data.dat
```

```
147 0:3 1:4 2:1 3:8 4:7 5:4 6:34 7:9 8:8 9:17 10:220 11:11 12:7 13:1 14:2 15:52 16:12 17:5
19:2 20:1 21:59 22:21 23:8 24:3 25:4 26:7 27:7 28:39 29:2 30:20 32:1 33:21 34:17 35:6 36:4
37:1 38:11 40:81 41:1 42:3 43:1 44:5 45:1 46:17 47:44 48:11 49:57 50:36 52:11 53:6 54:43
55:3 56:23 57:3 58:9 59:14 60:47 61:11 62:14 63:53 64:16 66:8 69:3 70:2 71:31 72:20 73:3
74:2 75:80 76:6 77:4 78:3 79:7 80:7 81:24 82:4 83:7 84:1 85:6 86:1 87:1 88:10 89:2 90:13
91:8 92:9 93:4 94:11 95:47 97:9 98:7 99:5 100:1 101:2 102:27 104:4 105:8 106:6 107:15 108:1
109:3 110:4 111:35 112:8 113:64 114:8 115:41 116:1 117:9 118:20 119:197 120:2 121:13 122:3
123:5 125:26 126:14 127:10 128:2 129:10 130:16 131:23 132:6 134:4 135:15 136:16 137:34
138:5 139:3 140:9 141:8 142:1 143:2 144:4 145:4 146:6 147:4 148:83 149:9 150:15 151:24
152:18 153:4 154:1 155:9 156:13 157:32

157 0:6 1:6 2:3 3:7 4:7 5:9 6:39 7:12 8:17 9:29 10:12 11:26 12:1 13:7 14:14 15:32 16:9 17:3
18:3 19:25 20:5 21:23 22:11 23:7 24:10 25:10 26:4 27:20 28:31 29:11 30:24 31:10 32:4 33:20
34:29 35:6 36:15 37:7 38:7 39:1 40:20 41:21 42:4 43:8 44:7 45:12 46:7 47:41 48:8 49:35 50:6
51:1 52:13 53:16 54:40 55:6 56:47 57:6 58:6 59:49 60:11 61:13 62:34 63:56 64:57 65:4 66:4
67:4 68:8 69:11 70:9 71:37 72:16 73:17 74:10 75:24 76:8 77:9 78:9 79:5 80:16 81:20 82:8
83:21 84:12 85:13 86:4 87:7 88:16 89:4 90:119 91:10 92:27 93:16 94:11 95:11 97:16 98:15
99:3 100:2 101:6 102:14 103:6 104:7 105:16 106:23 107:20 108:1 109:2 110:14 111:16 112:4
113:8 114:8 115:4 116:7 117:14 118:23 119:23 120:17 121:7 122:8 123:7 124:14 125:10 126:45
127:10 128:1 129:7 130:24 131:42 132:14 133:1 134:9 135:8 136:17 137:11 138:29 139:6 140:24
141:22 142:5 143:4 144:5 145:10 146:17 147:4 148:15 149:27 150:18 151:38 152:49 153:7 154:9
```

```
155:21 156:6 157:45
```

From the README:

[label] is a file where each line is the corresponding label for [data]. The labels must be 0, 1, ..., C-1, if we have C classes.

Sample training label file.

In [8]:
```
!head -n 2 Wang/sample_data/Labelme/images/train-label.dat
```

```
0

0
```

Estimate parameters

In [14]:
```
!slda est Wang/sample_data/Labelme/images/train-data.dat Wang/sample_data/Labelme/images/tra
```

```
reading data from Wang/sample_data/Labelme/images/train-data.dat
number of docs  : 800
number of terms : 158
number of total words : 1920800

reading labels from Wang/sample_data/Labelme/images/train-label.dat
number of classes : 8

alpha is esimated ...
var max iter 32767
var convergence 7.71E+11
em max iter 32767
em convergence 7.71E+11
L2 penalty 0.00E+00
number of topics is 10
models will be saved in slda_out
initializing ...
**** em iteration 1 ****
**** e-step ****
document 0
document 100
document 200
document 300
document 400
document 500
document 600
document 700
likelihood: 0.0000000000
**** m-step ****
maximizing ...
final f: 0.000000
**** em iteration 2 ****
**** e-step ****
document 0
document 100
document 200
document 300
document 400
document 500
document 600
document 700
likelihood: 0.0000000000
**** m-step ****
```

```
maximizing ...
final f: 0.000000
**** em iteration 3 ****
**** e-step ****
document 0
document 100
document 200
document 300
document 400
document 500
document 600
document 700
likelihood: 0.0000000000
**** m-step ****
maximizing ...
final f: 0.000000
final e step document 0
final e step document 100
final e step document 200
final e step document 300
final e step document 400
final e step document 500
final e step document 600
final e step document 700
```

In [15]:
```
!head slda_out/final.model.text
```

```
alpha: 0.500000
number of topics: 10
size of vocab: 158
number of classes: 8
betas:
-5.488458 -4.922297 -5.021655 -5.415970 -5.034608 -5.229130 -4.941934 -4.789785 -5.075424
-4.976122 -5.243023 -4.726343 -5.425503 -5.528802 -5.113379 -4.913542 -5.073925 -5.336739
-5.182619 -5.383589 -5.301550 -4.833821 -4.880092 -4.863508 -5.168986 -4.901987 -5.280771
-5.345651 -4.977028 -5.124606 -5.211888 -5.266146 -5.213417 -4.765597 -5.270086 -5.017243
-5.163784 -5.454535 -4.807282 -5.285283 -4.982480 -5.303223 -5.313002 -4.967558 -4.958919
-5.333174 -4.776536 -4.749650 -5.299568 -4.158433 -5.844539 -5.511257 -4.650704 -4.671015
-4.941060 -4.469864 -5.108114 -4.822241 -5.150853 -4.988954 -5.275669 -4.694190 -5.384610
-4.869186 -4.947197 -5.560640 -5.127059 -5.196629 -5.094349 -5.357732 -5.065145 -4.834738
-4.998321 -4.937134 -5.022208 -4.969880 -5.025928 -4.826389 -5.011052 -4.613482 -5.015907
-5.090535 -4.990178 -5.094433 -5.156439 -4.869457 -5.029502 -5.177439 -5.209884 -5.347071
-4.658691 -4.818239 -5.214565 -5.299568 -4.930697 -5.562537 -4.747066 -4.933946 -5.345978
-4.824767 -5.738370 -5.764414 -5.080017 -5.180210 -5.016928 -5.438236 -5.117107 -4.958771
-5.572486 -5.219653 -5.254916 -5.069194 -5.254020 -5.243812 -4.803085 -5.280771 -5.059307
-5.161693 -4.835655 -4.828600 -5.200870 -5.340424 -5.578531 -5.907617 -5.544262 -5.469621
-4.911914 -4.978918 -5.279544 -5.325228 -4.908312 -4.990713 -4.947417 -6.304881 -4.991786
-5.167432 -5.104848 -4.779382 -4.942955 -5.143602 -5.111909 -5.401090 -4.933802 -5.721854
-4.835328 -4.922583 -5.132866 -4.563615 -5.131015 -5.119629 -4.734368 -4.974765 -4.920940
-5.072678 -4.632134 -5.227577 -4.991939 -4.910005
-5.488458 -4.922297 -5.021655 -5.415970 -5.034608 -5.229130 -4.941934 -4.789785 -5.075424
-4.976122 -5.243023 -4.726343 -5.425503 -5.528802 -5.113379 -4.913542 -5.073925 -5.336739
-5.182619 -5.383589 -5.301550 -4.833821 -4.880092 -4.863508 -5.168986 -4.901987 -5.280771
-5.345651 -4.977028 -5.124606 -5.211888 -5.266146 -5.213417 -4.765597 -5.270086 -5.017243
-5.163784 -5.454535 -4.807282 -5.285283 -4.982480 -5.303223 -5.313002 -4.967558 -4.958919
-5.333174 -4.776536 -4.749650 -5.299568 -4.158433 -5.844539 -5.511257 -4.650704 -4.671015
-4.941060 -4.469864 -5.108114 -4.822241 -5.150853 -4.988954 -5.275669 -4.694190 -5.384610
-4.869186 -4.947197 -5.560640 -5.127059 -5.196629 -5.094349 -5.357732 -5.065145 -4.834738
-4.998321 -4.937134 -5.022208 -4.969880 -5.025928 -4.826389 -5.011052 -4.613482 -5.015907
-5.090535 -4.990178 -5.094433 -5.156439 -4.869457 -5.029502 -5.177439 -5.209884 -5.347071
-4.658691 -4.818239 -5.214565 -5.299568 -4.930697 -5.562537 -4.747066 -4.933946 -5.345978
-4.824767 -5.738370 -5.764414 -5.080017 -5.180210 -5.016928 -5.438236 -5.117107 -4.958771
-5.572486 -5.219653 -5.254916 -5.069194 -5.254020 -5.243812 -4.803085 -5.280771 -5.059307
-5.161693 -4.835655 -4.828600 -5.200870 -5.340424 -5.578531 -5.907617 -5.544262 -5.469621
-4.911914 -4.978918 -5.279544 -5.325228 -4.908312 -4.990713 -4.947417 -6.304881 -4.991786
```

```
-5.167432 -5.104848 -4.779382 -4.942955 -5.143602 -5.111909 -5.401090 -4.933802 -5.721854
-4.835328 -4.922583 -5.132866 -4.563615 -5.131015 -5.119629 -4.734368 -4.974765 -4.920940
-5.072678 -4.632134 -5.227577 -4.991939 -4.910005
-5.488458 -4.922297 -5.021655 -5.415970 -5.034608 -5.229130 -4.941934 -4.789785 -5.075424
-4.976122 -5.243023 -4.726343 -5.425503 -5.528802 -5.113379 -4.913542 -5.073925 -5.336739
-5.182619 -5.383589 -5.301550 -4.833821 -4.880092 -4.863508 -5.168986 -4.901987 -5.280771
-5.345651 -4.977028 -5.124606 -5.211888 -5.266146 -5.213417 -4.765597 -5.270086 -5.017243
-5.163784 -5.454535 -4.807282 -5.285283 -4.982480 -5.303223 -5.313002 -4.967558 -4.958919
-5.333174 -4.776536 -4.749650 -5.299568 -4.158433 -5.844539 -5.511257 -4.650704 -4.671015
-4.941060 -4.469864 -5.108114 -4.822241 -5.150853 -4.988954 -5.275669 -4.694190 -5.384610
-4.869186 -4.947197 -5.560640 -5.127059 -5.196629 -5.094349 -5.357732 -5.065145 -4.834738
-4.998321 -4.937134 -5.022208 -4.969880 -5.025928 -4.826389 -5.011052 -4.613482 -5.015907
-5.090535 -4.990178 -5.094433 -5.156439 -4.869457 -5.029502 -5.177439 -5.209884 -5.347071
-4.658691 -4.818239 -5.214565 -5.299568 -4.930697 -5.562537 -4.747066 -4.933946 -5.345978
-4.824767 -5.738370 -5.764414 -5.080017 -5.180210 -5.016928 -5.438236 -5.117107 -4.958771
-5.572486 -5.219653 -5.254916 -5.069194 -5.254020 -5.243812 -4.803085 -5.280771 -5.059307
-5.161693 -4.835655 -4.828600 -5.200870 -5.340424 -5.578531 -5.907617 -5.544262 -5.469621
-4.911914 -4.978918 -5.279544 -5.325228 -4.908312 -4.990713 -4.947417 -6.304881 -4.991786
-5.167432 -5.104848 -4.779382 -4.942955 -5.143602 -5.111909 -5.401090 -4.933802 -5.721854
-4.835328 -4.922583 -5.132866 -4.563615 -5.131015 -5.119629 -4.734368 -4.974765 -4.920940
-5.072678 -4.632134 -5.227577 -4.991939 -4.910005
-5.488458 -4.922297 -5.021655 -5.415970 -5.034608 -5.229130 -4.941934 -4.789785 -5.075424
-4.976122 -5.243023 -4.726343 -5.425503 -5.528802 -5.113379 -4.913542 -5.073925 -5.336739
-5.182619 -5.383589 -5.301550 -4.833821 -4.880092 -4.863508 -5.168986 -4.901987 -5.280771
-5.345651 -4.977028 -5.124606 -5.211888 -5.266146 -5.213417 -4.765597 -5.270086 -5.017243
-5.163784 -5.454535 -4.807282 -5.285283 -4.982480 -5.303223 -5.313002 -4.967558 -4.958919
-5.333174 -4.776536 -4.749650 -5.299568 -4.158433 -5.844539 -5.511257 -4.650704 -4.671015
-4.941060 -4.469864 -5.108114 -4.822241 -5.150853 -4.988954 -5.275669 -4.694190 -5.384610
-4.869186 -4.947197 -5.560640 -5.127059 -5.196629 -5.094349 -5.357732 -5.065145 -4.834738
-4.998321 -4.937134 -5.022208 -4.969880 -5.025928 -4.826389 -5.011052 -4.613482 -5.015907
-5.090535 -4.990178 -5.094433 -5.156439 -4.869457 -5.029502 -5.177439 -5.209884 -5.347071
-4.658691 -4.818239 -5.214565 -5.299568 -4.930697 -5.562537 -4.747066 -4.933946 -5.345978
-4.824767 -5.738370 -5.764414 -5.080017 -5.180210 -5.016928 -5.438236 -5.117107 -4.958771
-5.572486 -5.219653 -5.254916 -5.069194 -5.254020 -5.243812 -4.803085 -5.280771 -5.059307
-5.161693 -4.835655 -4.828600 -5.200870 -5.340424 -5.578531 -5.907617 -5.544262 -5.469621
-4.911914 -4.978918 -5.279544 -5.325228 -4.908312 -4.990713 -4.947417 -6.304881 -4.991786
-5.167432 -5.104848 -4.779382 -4.942955 -5.143602 -5.111909 -5.401090 -4.933802 -5.721854
-4.835328 -4.922583 -5.132866 -4.563615 -5.131015 -5.119629 -4.734368 -4.974765 -4.920940
-5.072678 -4.632134 -5.227577 -4.991939 -4.910005
-5.488458 -4.922297 -5.021655 -5.415970 -5.034608 -5.229130 -4.941934 -4.789785 -5.075424
-4.976122 -5.243023 -4.726343 -5.425503 -5.528802 -5.113379 -4.913542 -5.073925 -5.336739
-5.182619 -5.383589 -5.301550 -4.833821 -4.880092 -4.863508 -5.168986 -4.901987 -5.280771
-5.345651 -4.977028 -5.124606 -5.211888 -5.266146 -5.213417 -4.765597 -5.270086 -5.017243
-5.163784 -5.454535 -4.807282 -5.285283 -4.982480 -5.303223 -5.313002 -4.967558 -4.958919
-5.333174 -4.776536 -4.749650 -5.299568 -4.158433 -5.844539 -5.511257 -4.650704 -4.671015
-4.941060 -4.469864 -5.108114 -4.822241 -5.150853 -4.988954 -5.275669 -4.694190 -5.384610
-4.869186 -4.947197 -5.560640 -5.127059 -5.196629 -5.094349 -5.357732 -5.065145 -4.834738
-4.998321 -4.937134 -5.022208 -4.969880 -5.025928 -4.826389 -5.011052 -4.613482 -5.015907
-5.090535 -4.990178 -5.094433 -5.156439 -4.869457 -5.029502 -5.177439 -5.209884 -5.347071
-4.658691 -4.818239 -5.214565 -5.299568 -4.930697 -5.562537 -4.747066 -4.933946 -5.345978
-4.824767 -5.738370 -5.764414 -5.080017 -5.180210 -5.016928 -5.438236 -5.117107 -4.958771
-5.572486 -5.219653 -5.254916 -5.069194 -5.254020 -5.243812 -4.803085 -5.280771 -5.059307
-5.161693 -4.835655 -4.828600 -5.200870 -5.340424 -5.578531 -5.907617 -5.544262 -5.469621
-4.911914 -4.978918 -5.279544 -5.325228 -4.908312 -4.990713 -4.947417 -6.304881 -4.991786
-5.167432 -5.104848 -4.779382 -4.942955 -5.143602 -5.111909 -5.401090 -4.933802 -5.721854
-4.835328 -4.922583 -5.132866 -4.563615 -5.131015 -5.119629 -4.734368 -4.974765 -4.920940
-5.072678 -4.632134 -5.227577 -4.991939 -4.910005
```

In [ ]: