

Figure 4 generated as in the Lewis paper (using raw data with 1151 features):

- the data is clustered to 3 clusters
- the mean is found for each cluster and plotted

Figure 2 generated using fastmapped data of 4 features. The values of each feature are plotted.

-----  
-----CBR vs FSS+CBR-----  
-----

	Before		After	
Breast Cancer	pd	pf	pd	pf
CBR	61	40	60	40
FSS+CBR	56	44	70	30

-----

	Before		After	
Heart	pd	pf	pd	pf
CBR	22	12	0	4
FSS+CBR	33	13	38	10

-----

	Before		After	
Lymph	pd	pf	pd	pf
CBR	70	0	59	0
FSS+CBR	60	8	50	3

-----

	Before		After	
Pima Diabetes	pd	pf	pd	pf
CBR	58	42	61	39
FSS+CBR	58	42	58	42

-----  
-----Experiment 1 - BestK-----  
-----

Dataset	bestK
Breast Cancer	3
Heart	1
Lymph	1
Pima Diabetes	3

-----

Naive Bayes

Dataset	pd	pf
Breast Cancer	100	0
Heart	22	8
Lymph	100	0
Pima Diabetes	100	0

-----  
---Experiment 2 - Within System Brittle Check-----  
-----

NB. Definition of brittleness  
if NUN >= NLN  
then brittle = yes

Before	After
<pre> ----- Dataset          pd  pf  brittle? ----- Breast Cancer   61  40  yes (7, 20%)  #key, ties, win, loss, win-loss NUN,    0,  1,  0,    1 NLN,    0,  0,  1,   -1 ----- </pre>	<pre> ----- Dataset          pd  pf  brittle? ----- Breast Cancer   60  40  no (7, 20%)  #key, ties, win, loss, win-loss NLN,    0,  1,  0,    1 NUN,    0,  0,  1,   -1 ----- </pre>
<pre> ----- Dataset          pd  pf  brittle? ----- Heart           22  12  yes (9, 12%)  #key, ties, win, loss, win-loss NUN,    1,  0,  0,    0 NLN,    1,  0,  0,    0 ----- </pre>	<pre> ----- Dataset          pd  pf  brittle? ----- Heart           0   4  yes (9, 12%)  #key, ties, win, loss, win-loss NLN,    1,  0,  0,    0 NUN,    1,  0,  0,    0 ----- </pre>
<pre> ----- Dataset          pd  pf  brittle? ----- Lymph           70   0  yes (17, 14%)  #key, ties, win, loss, win-loss NUN,    1,  0,  0,    0 NLN,    1,  0,  0,    0 ----- </pre>	<pre> ----- Dataset          pd  pf  brittle? ----- Lymph           59   0  no (17, 14%)  #key, ties, win, loss, win-loss NLN,    0,  1,  0,    1 NUN,    0,  0,  1,   -1 ----- </pre>
<pre> ----- Dataset          pd  pf  brittle? ----- Pima Diabetes   58  42  yes (2, 12%)  #key, ties, win, loss, win-loss NUN,    1,  0,  0,    0 NLN,    1,  0,  0,    0 ----- </pre>	<pre> ----- Dataset          pd  pf  brittle? ----- Pima Diabetes   61  39  yes (2, 12%)  #key, ties, win, loss, win-loss NUN,    1,  0,  0,    0 NLN,    1,  0,  0,    0 ----- </pre>

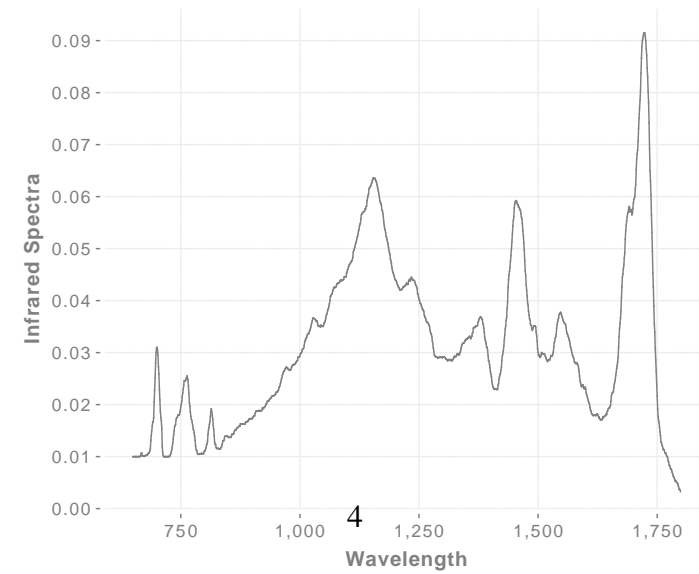
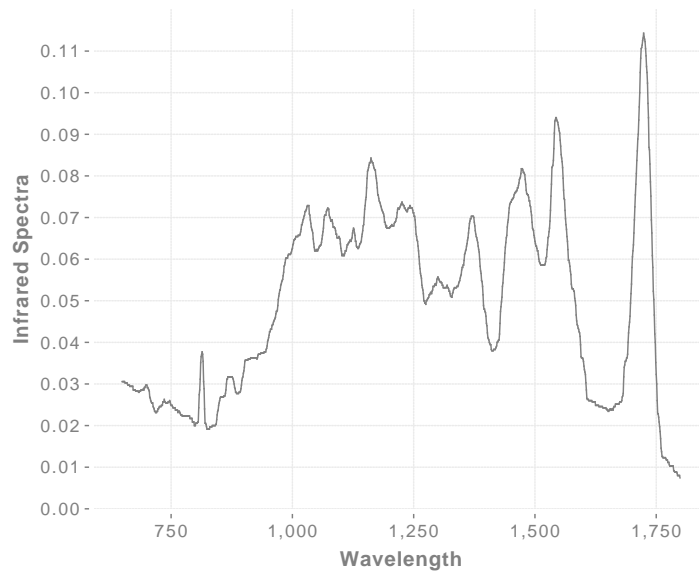
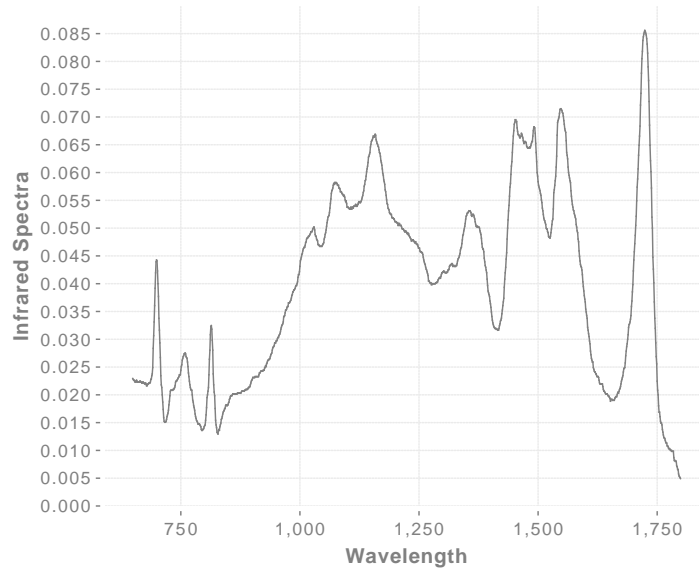


Figure 1: Infrared spectra of the mean of classes produced by kmeans clustering on 185 infrared spectra from 37 vehicles.

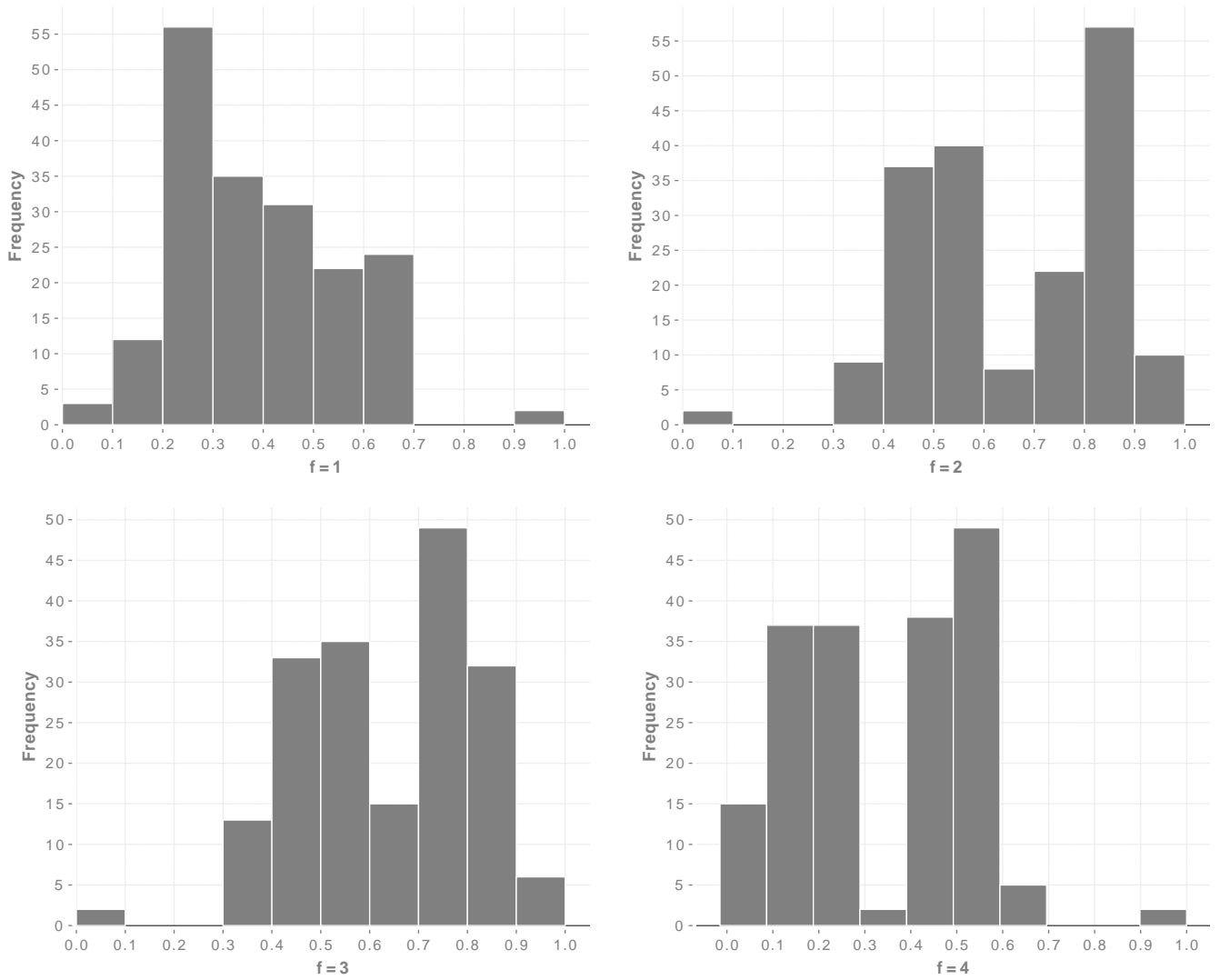


Figure 2: FastMapped infrared spectra for 4 features.

Clusters	n	Final Instances	Reduction%
bc (286, 10, 2)	n=1	272	95
	n=2	272	95
	n=4	200	70
	<b>n=7</b>	58	<b>20</b>
	n=8	8	3
	n=16	0	0
heart (297, 14, 5)	n=1	249	84
	n=2	200	67
	n=4	176	59
	n=8	64	22
	<b>n=9</b>	35	<b>12</b>
	n=16	0	0
lym (148, 19, 4)	n=1	120	81
	n=2	118	80
	n=4	115	78
	n=8	115	78
	n=16	32	22
	<b>n=17</b>	20	<b>14</b>
pima (768, 9, 2)	n=1	375	49
	<b>n=2</b>	95	<b>12</b>
	n=4	13	2
	n=8	0	0
	n=16	0	0
tumor (339, 18, 21)	n=1	270	80
	n=2	255	75
	n=4	255	75
	n=8	255	75
	n=16	193	57
	<b>n=18</b>	126	<b>37</b>

Figure 3: Instance selection using the CLIFF selector. The Reduction% column shows the percentage of the original data set left after selection.

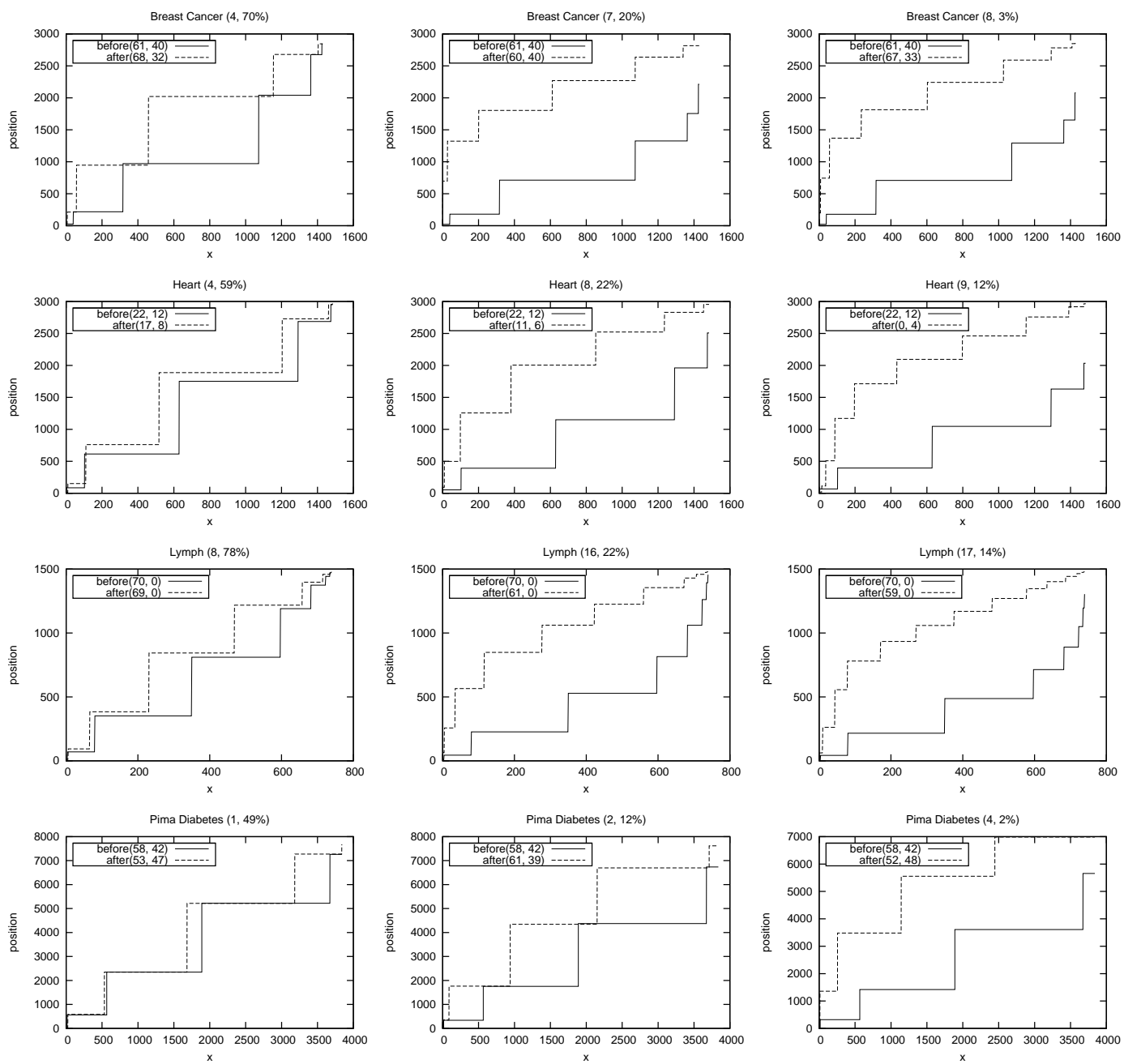


Figure 4: Breast Cancer, Heart, Lymph, Pima.

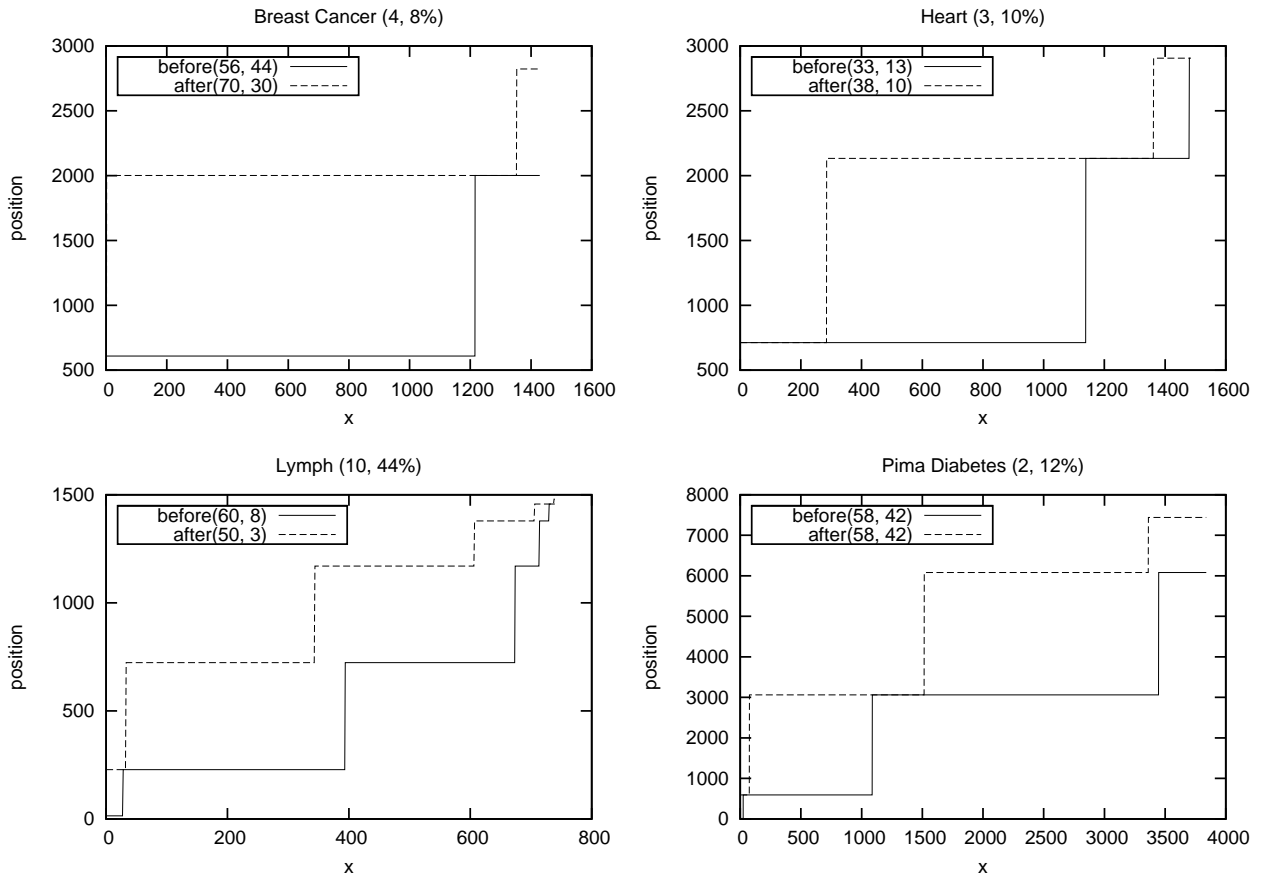


Figure 5: Breast Cancer, Heart, Lymph, Pima.

Clusters	n	Final Instances	Reduction%
bc (286, 10, 2)	n=1	55	19
	n=2	55	19
	n=3	50	17
	<b>n=4</b>	24	<b>8</b>
heart (297, 14, 5)	n=1	79	27
	n=2	52	18
	<b>n=3</b>	30	<b>10</b>
	n=4	22	7
lym (148, 19, 4)	n=10	65	44
pima (768, 9, 2)	n=1	362	47
	<b>n=2</b>	92	<b>12</b>
	n=3	36	5
	n=4	11	1

Figure 6: Instance selection using the CLIFF selector. The Reduction% column shows the percentage of the original data set left after selection.