Angela Krontiris
CISC 6930: Data Mining
Assignment 3
Due: November 2

3) Filter Method

a) List the features from highest |r| (the absolute value of r) to lowest, along with their |r| values. Why would one be interested in the absolute value of r rather than the raw value?

Highest accuracy generated by ranked features by correlation, r:

18 [0.925531914893617]
16 [0.9243498817966903]
17 [0.91725768321513]
19 [0.9113475177304965]
15 [0.9101654846335697]
13 [0.9089834515366431]
10 [0.9042553191489362]
20 [0.9030732860520094]
9 [0.9018912529550828]
14 [0.8971631205673759]
21 [0.8947990543735225]
7 [0.8947990543735225]
12 [0.8912529550827423]
11 [0.8888888888888888]
8 [0.8853427895981087]
22 [0.8841607565011821]
24 [0.8794326241134752]
23 [0.8794326241134752]
6 [0.8747044917257684]
25 [0.8747044917257684]
26 [0.8699763593380615]
5 [0.8617021276595744]
27 [0.8557919621749409]
28 [0.8498817966903073]
4 [0.8368794326241135]
29 [0.8368794326241135]

3 [0.8356973995271868]
30 [0.8356973995271868]
2 [0.8309692671394799]
34 [0.8262411347517731]
32 [0.8262411347517731]
31 [0.8250591016548463]
1 [0.8226950354609929]
33 [0.8096926713947991]
0 [0.7978723404255319]

Taking the absolute value of the correlation coefficient (r) measures the strength of the relationship. In order to rank features by their correlation to the class function, we have to take the absolute value of r. We are interested in the magnitude of the correlation without regard to direction.

(b) Select the features that have the highest m values of |r|, and run LOOCV on the dataset restricted to only those m features. Which value of m gives the highest LOOCV classification accuracy, and what is the value of this optimal accuracy?

These are the top features using filter method:
[4, 13, 14, 16, 7, 22, 26, 1, 20, 31, 34, 2, 28, 25, 19, 17, 32, 8, 0, 10]

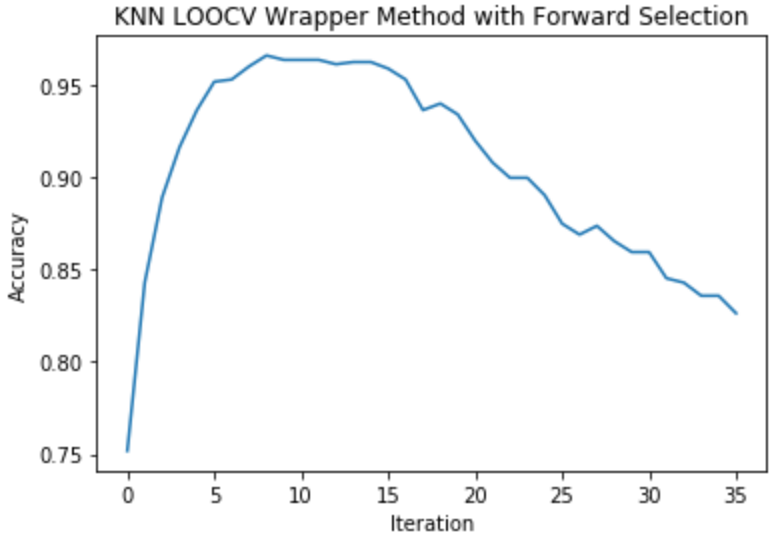
These top 20 correlated features yield an accuracy of 0.925531914893617

4) Wrapper Method
(a) Show the set of selected features at each step, as it grows from size zero to its final size (increasing in size by exactly one feature at each step).

Top features selected at each step:
[['f20'], ['f10', 'f20'], ['f19', 'f20', 'f10'], ['f8', 'f20', 'f10', 'f19'], ['f7', 'f20', 'f10', 'f19', 'f8'], ['f14', 'f20', 'f10', 'f19', 'f8', 'f7'], ['f2', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14'], ['f1', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2'], ['f16', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1'], ['f4', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16'], ['f13', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4'], ['f22', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13'], ['f26', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22'], ['f34', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26'], ['f25', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34'], ['f32', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25'], ['f18', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4',

'f13', 'f22', 'f26', 'f34', 'f25', 'f32'], ['f31', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18'], ['f28', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31'], ['f24', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28'], ['f35', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24'], ['f6', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35'], ['f3', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6'], ['f15', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3'], ['f29', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15'], ['f11', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29'], ['f9', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29', 'f11'], ['f23', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29', 'f11', 'f9'], ['f17', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29', 'f11', 'f9', 'f23'], ['f33', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29', 'f11', 'f9', 'f23', 'f17'], ['f0', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29', 'f11', 'f9', 'f23', 'f17', 'f33'], ['f5', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29', 'f11', 'f9', 'f23', 'f17', 'f33', 'f0'], ['f21', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29', 'f11', 'f9', 'f23', 'f17', 'f33', 'f0', 'f5'], ['f30', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29', 'f11', 'f9', 'f23', 'f17', 'f33', 'f0', 'f5', 'f21'], ['f12', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29', 'f11', 'f9', 'f23', 'f17', 'f33', 'f0', 'f5', 'f21', 'f30'], ['f27', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1', 'f16', 'f4', 'f13', 'f22', 'f26', 'f34', 'f25', 'f32', 'f18', 'f31', 'f28', 'f24', 'f35', 'f6', 'f3', 'f15', 'f29', 'f11', 'f9', 'f23', 'f17', 'f33', 'f0', 'f5', 'f21', 'f30', 'f12']]

(b) What is the LOOCV accuracy over the final set of selected features?



The maximum accuracy is attained for the KNN LOOCV Wrapper Method with Forward Selection is: **0.9657210401891253**

The maximum accuracy occurs on the 8 iteration

The feature combination that yielded the highest accuracy is:
['f16', 'f20', 'f10', 'f19', 'f8', 'f7', 'f14', 'f2', 'f1']