Cast as a classification problem, the k-means clustering algorithm achieves a sensitivity of 0.613, a specificity of 0.997, and an F1 score of 0.758. The confusion matrix is as follows.

```
     -------------------------- Diagnosis --------------------------
     |
     |          tp = 130                            fp = 1
 Prediction
  (cluster)
     |          fn = 82                             tn = 356
     |
     ---------------------------------------------------------------
```

Essentially, k-means does an excellent job not classifying benign patients as malignant, but it fails to classify many of the malignant patients as such.

Appendix

The following Matlab script performs the k-means clustering.

run-kmeans.m

```
x = dlmread('wdbc-values.data', ',');
c = kmeans(x, 2);
dlmwrite('wdbc-clusters.data', c);
exit
```

The following shell script can be used to download the data, process it, run the Matlab script, and compare the clustering results with actual diagnoses.

run-all.sh

```
#!/bin/bash
DATASERVER=http://archive.ics.uci.edu
DATAPATH=ml/machine-learning-databases/breast-cancer-wisconsin
curl -o wdbc.data $DATASERVER/$DATAPATH/wdbc.data
perl -ne '@f = split/,/; print(join(",", @f[2..31]))' < wdbc.data > wdbc-values.data
perl -ne '@f = split/,/; printf("%s\n", $f[1])' < wdbc.data > wdbc-diagnoses.data
/Applications/MATLAB_R2011b.app/bin/matlab -nodisplay < run-kmeans.m
echo -e "\n\n\n======Results======"
paste -d: wdbc-clusters.data wdbc-diagnoses.data | sort | uniq -c
```

Running the shell script on my desktop gives the following terminal output.

```
dhrasmus:hw04 standage$ bash run-all.sh
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  121k  100  121k    0     0   131k      0 --:--:-- --:--:-- --:--:--  253k

                            < M A T L A B (R) >
                  Copyright 1984-2011 The MathWorks, Inc.
                    R2011b (7.13.0.564) 64-bit (maci64)
```

```
                              August 13, 2011


To get started, type one of these: helpwin, helpdesk, or demo.
For product information, visit www.mathworks.com.

>> >> >> >>


======Results======
   1 1:B
 130 1:M
 356 2:B
  82 2:M
```