Data Mining Lab 2
Michael Francki
500554567

————————————————2————————————————

a. Cross-validation folds 10

Correctly Classified Instances	42	73.6842 %
Incorrectly Classified Instances	15	26.3158 %

b. Cross-validation folds 5

Correctly Classified Instances	44	77.193 %
Incorrectly Classified Instances	13	22.807 %

c. Cross-validation folds 2

Correctly Classified Instances	41	71.9298 %
Incorrectly Classified Instances	16	28.0702 %

d. Percentage split 66%

Correctly Classified Instances	17	89.4737 %
Incorrectly Classified Instances	2	10.5263 %

e. Percentage split 33%

Correctly Classified Instances	31	81.5789 %
Incorrectly Classified Instances	7	18.4211

f. Use training set

Correctly Classified Instances	50	87.7193 %
Incorrectly Classified Instances	7	12.2807 %

d, e and f method seem to be have good accuracy but have most likely over fit
the data. The cross validation methods seem to be worse but most likely when
given fresh data will out perform the others method.

————————————————3————————————————

————————————————4————————————————

a) Training set

=== Evaluation on training set ===

```
Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correctly Classified Instances         147               98      %
Incorrectly Classified Instances         3                2      %
Kappa statistic                          0.97
Mean absolute error                      0.0233
Root mean squared error                  0.108
Relative absolute error                  5.2482 %
Root relative squared error             22.9089 %
Total Number of Instances              150

b)Cross Validation
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         144               96      %
Incorrectly Classified Instances         6                4      %
Kappa statistic                          0.94
Mean absolute error                      0.035
Root mean squared error                  0.1586
Relative absolute error                  7.8705 %
Root relative squared error             33.6353 %
Total Number of Instances              150


a) 98%
b) 96%

b) is more realistic  using cross validation  will help correct the
overfitting


————————————————————5————————————————————

the errors are located when petal width > 0.6

————————————————————6————————————————————

 simple Linear Regression would not run
```