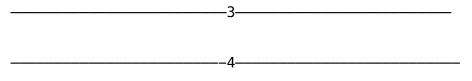
| | | _ |
|---|----|-----------|
| a. Cross-validation folds 10 Correctly Classified Instances | 42 | 73.6842 % |
| Incorrectly Classified Instances | 15 | 26.3158 % |
| <pre>b. Cross-validation folds 5 Correctly Classified Instances</pre> | 44 | 77.193 % |
| Incorrectly Classified Instances | 13 | 22.807 % |
| <pre>c. Cross-validation folds 2 Correctly Classified Instances</pre> | 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 |
| <pre>f. Use training set Correctly Classified Instances</pre> | 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.



a) Training set

=== Evaluation on training set ===

| - · | | | | | | | | | |
|----------------|-------|----|------|-------|----|----------|-------|------|---------|
| lime | taken | to | test | model | on | training | data: | 0.01 | seconds |

=== Summary ===

| Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error Relative absolute error Root relative squared error Total Number of Instances | 147 3 0.97 0.0233 0.108 5.2482 % 22.9089 % 150 | 98 2 | % % % |
|---|---|---------|----------|
| <pre>b)Cross Validation === Stratified cross-validation === === Summary ===</pre> | | | |
| Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error Relative absolute error Root relative squared error Total Number of Instances | 144 6 0.94 0.035 0.1586 7.8705 % 33.6353 % 150 | 96 4 | % % |

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

simple Linear Regression would not run