

Wheat Classification

January 2, 2022

```
In [2]: wheat <- read.csv("wheat.csv", header = T)
wheat
```

	area <dbl>	perimeter <dbl>	compactness <dbl>	length <dbl>	width <dbl>	asymmetry <dbl>	groove <dbl>	type <fct>
	15.26	14.84	0.8710	5.763	3.312	2.2210	5.220	A
	14.88	14.57	0.8811	5.554	3.333	1.0180	4.956	A
	14.29	14.09	0.9050	5.291	3.337	2.6990	4.825	A
	13.84	13.94	0.8955	5.324	3.379	2.2590	4.805	A
	16.14	14.99	0.9034	5.658	3.562	1.3550	5.175	A
	14.38	14.21	0.8951	5.386	3.312	2.4620	4.956	A
	14.69	14.49	0.8799	5.563	3.259	3.5860	5.219	A
	14.11	14.10	0.8911	5.420	3.302	2.7000	5.000	A
	16.63	15.46	0.8747	6.053	3.465	2.0400	5.877	A
	16.44	15.25	0.8880	5.884	3.505	1.9690	5.533	A
	15.26	14.85	0.8696	5.714	3.242	4.5430	5.314	A
	14.03	14.16	0.8796	5.438	3.201	1.7170	5.001	A
	13.89	14.02	0.8880	5.439	3.199	3.9860	4.738	A
	13.78	14.06	0.8759	5.479	3.156	3.1360	4.872	A
	13.74	14.05	0.8744	5.482	3.114	2.9320	4.825	A
	14.59	14.28	0.8993	5.351	3.333	4.1850	4.781	A
	13.99	13.83	0.9183	5.119	3.383	5.2340	4.781	A
	15.69	14.75	0.9058	5.527	3.514	1.5990	5.046	A
	14.70	14.21	0.9153	5.205	3.466	1.7670	4.649	A
	12.72	13.57	0.8686	5.226	3.049	4.1020	4.914	A
	14.16	14.40	0.8584	5.658	3.129	3.0720	5.176	A
	14.11	14.26	0.8722	5.520	3.168	2.6880	5.219	A
	15.88	14.90	0.8988	5.618	3.507	0.7651	5.091	A
	12.08	13.23	0.8664	5.099	2.936	1.4150	4.961	A
	15.01	14.76	0.8657	5.789	3.245	1.7910	5.001	A
	16.19	15.16	0.8849	5.833	3.421	0.9030	5.307	A
	13.02	13.76	0.8641	5.395	3.026	3.3730	4.825	A
	12.74	13.67	0.8564	5.395	2.956	2.5040	4.869	A
	14.11	14.18	0.8820	5.541	3.221	2.7540	5.038	A
A data.frame: 200 CE 8	13.45	14.02	0.8604	5.516	3.065	3.5310	5.097	A
	11.48	13.05	0.8473	5.180	2.758	5.876	5.002	C
	12.21	13.47	0.8453	5.357	2.893	1.661	5.178	C
	11.41	12.95	0.8560	5.090	2.775	4.957	4.825	C
	12.46	13.41	0.8706	5.236	3.017	4.987	5.147	C
	12.19	13.36	0.8579	5.240	2.909	4.857	5.158	C
	11.65	13.07	0.8575	5.108	2.850	5.209	5.135	C
	12.89	13.77	0.8541	5.495	3.026	6.185	5.316	C
	11.81	13.45	0.8198	5.413	2.716	4.898	5.352	C
	10.91	12.80	0.8372	5.088	2.675	4.179	4.956	C
	11.23	12.82	0.8594	5.089	2.821	7.524	4.957	C
	10.59	12.41	0.8648	4.899	2.787	4.975	4.794	C
	10.93	12.80	0.8390	5.046	2.717	5.398	5.045	C
	11.27	12.86	0.8563	5.091	2.804	3.985	5.001	C
	11.87	13.02	0.8795	5.132	2.953	3.597	5.132	C
	12.11	13.27	0.8639	5.236	2.975	4.132	5.012	C
	12.80	13.47	0.8860	5.160	3.126	4.873	4.914	C
	12.79	13.53	0.8786	5.224	3.054	5.483	4.958	C
	13.37	13.78	0.8849	5.320	3.128	4.670	5.091	C
	12.62	13.67	0.8481	5.410	2.911	3.306	5.231	C
	12.76	13.38	0.8964	5.073	3.155	2.828	4.830	C

```
In [11]: options(repr.matrix.max.rows=600, repr.matrix.max.cols=200)
is.na(wheat)
#There are no missing values
```

[illegible]

```
In [22]: summary(wheat)
```

```
#The magnitudes and ranges of each column are quite different. Area column has  
#range ~10 whereas compactness column has range ~0.1
```

area	perimeter	compactness	length
Min. :10.59	Min. :12.41	Min. :0.8081	Min. :4.899
1st Qu.:12.38	1st Qu.:13.47	1st Qu.:0.8583	1st Qu.:5.257
Median :14.40	Median :14.38	Median :0.8745	Median :5.534
Mean :14.94	Mean :14.60	Mean :0.8721	Mean :5.639
3rd Qu.:17.41	3rd Qu.:15.78	3rd Qu.:0.8879	3rd Qu.:6.009
Max. :21.18	Max. :17.25	Max. :0.9183	Max. :6.675

width	asymmetry	groove	type
Min. :2.642	Min. :0.7651	Min. :4.519	A:68
1st Qu.:2.965	1st Qu.:2.4935	1st Qu.:5.043	B:69
Median :3.252	Median :3.5915	Median :5.226	C:63
Mean :3.273	Mean :3.6627	Mean :5.414	
3rd Qu.:3.564	3rd Qu.:4.7043	3rd Qu.:5.879	
Max. :4.033	Max. :8.4560	Max. :6.550	

```
In [38]: library(caret)
```

```
train_rows <- createDataPartition(y = wheat$type, p = 0.8, list = F)  
training <- wheat[train_rows, ]  
testing <- wheat[-train_rows, ]  
trctrl <- trainControl(method = "repeatedcv", number=10, repeats = 3)  
svm_linear <- train(type ~ ., data = training, method = "svmLinear", trControl = trctrl)  
svm_linear
```

Support Vector Machines with Linear Kernel

162 samples

7 predictor

3 classes: 'A', 'B', 'C'

Pre-processing: centered (7), scaled (7)

Resampling: Cross-Validated (10 fold, repeated 3 times)

Summary of sample sizes: 147, 145, 147, 146, 145, 145, ...

Resampling results:

Accuracy	Kappa
0.9151879	0.8727125

Tuning parameter 'C' was held constant at a value of 1

```
In [46]: testing_pred <- predict(svm_linear, newdata = testing)
```

```
confusionMatrix(testing_pred, testing$type)
```

```
#There's only one false prediction out of 68 rows of data and the accuracy
```

*#is 97%. The model is pretty accurate in associating 7 attributes with wheat type
#and predicting the wheat type correctly*

Confusion Matrix and Statistics

		Reference		
Prediction	A	B	C	
A	12	0	0	
B	0	13	0	
C	1	0	12	

Overall Statistics

Accuracy : 0.9737
 95% CI : (0.8619, 0.9993)
 No Information Rate : 0.3421
 P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.9605
 McNemar's Test P-Value : NA

Statistics by Class:

	Class: A	Class: B	Class: C
Sensitivity	0.9231	1.0000	1.0000
Specificity	1.0000	1.0000	0.9615
Pos Pred Value	1.0000	1.0000	0.9231
Neg Pred Value	0.9615	1.0000	1.0000
Prevalence	0.3421	0.3421	0.3158
Detection Rate	0.3158	0.3421	0.3158
Detection Prevalence	0.3158	0.3421	0.3421
Balanced Accuracy	0.9615	1.0000	0.9808

[1] "There's only one false prediction out of 68 rows of data and the accuracy \n is 97%. The r

```
In [41]: #classifications on unknown data
unknown <- read.csv("wheat-unknown.csv", header = T)
pred <- predict(svm_linear, newdata = unknown)
print("The predictions for the unknown wheat seeds are as follows, with 91.5% accuracy")
pred
```

[1] "The predictions for the unknown wheat seeds are as follows, with 91.5% accuracy:"

1. C 2. A 3. C 4. C 5. C 6. C 7. A 8. C 9. C 10. B
 Levels: 1. 'A' 2. 'B' 3. 'C'