



Chathura J Gunasekara

LETTER IMAGE RECOGNITION

UCI ML Repository

Recap from previous presentation

- ❖ 26 Classes, one for each letter A,B,C...ect and have even class distribution in the training set
- ❖ 16 features identified from letter images
- ❖ After preprocessing such as removing high correlated predictors dataset is left with 13 predictors.
- ❖ Checked for skewness and zero variance and not found any issues which can hinder a good classification performance
- ❖ Split the data set to 75% training and 25% testing sets.

Linear Classification Methods

- ❖ LDA
- ❖ PLSDA

1. Linear Discriminant Analysis

15000 samples

13 predictor

26 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L', 'M', 'N', 'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W', 'X', 'Y', 'Z'

Pre-processing: centered, scaled

Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 15000, 15000, 15000, 15000, 15000, 15000, ...

Resampling results across tuning parameters:

dimen	Accuracy	Kappa	Accuracy SD	Kappa SD
1	0.1685297	0.1355126	0.008434755	0.008705092
2	0.3497718	0.3236751	0.006712291	0.006945867
3	0.4289124	0.4060080	0.015855319	0.016455423
4	0.5390830	0.5205938	0.007712305	0.008013558
5	0.5833730	0.5666671	0.006934202	0.007205122
6	0.6151006	0.5996624	0.008165075	0.008487258
7	0.6250795	0.6100264	0.009064244	0.009421349
8	0.6726983	0.6595546	0.006922812	0.007192980
9	0.6816396	0.6688550	0.006811382	0.007074126
10	0.6904524	0.6780292	0.006649750	0.006907253
11	0.6867831	0.6742183	0.007103359	0.007377300
12	0.6925854	0.6802618	0.007167037	0.007444857
13	0.6915845	0.6792207	0.006972897	0.007243935

Kappa was used to select the optimal model using the largest value.
The final value used for the model was dimen = 12.

Confusion Matrix and Statistics for testing data

Overall Statistics

Accuracy : 0.6882

Kappa : 0.6757

2 . Partial Least Square Discrimination Analysis

15000 samples

13 predictor

26 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K',
'L', 'M', 'N', 'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W', 'X', 'Y', 'Z'

Pre-processing: centered, scaled

Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 15000, 15000, 15000, 15000, 15000, 15000, ...

Resampling results across tuning parameters:

ncomp	Accuracy	Kappa	Accuracy SD	Kappa SD
1	0.07420849	0.0372415	0.003271222	0.002700414
2	0.17892449	0.1469015	0.010474731	0.010607809

Kappa was used to select the optimal model using the largest value.
The final value used for the model was ncomp = 2.

Overall Statistics

Accuracy : 0.1908

Kappa : 0.1571

Non-Linear Classification methods

- ❖ `K-nn<-knn(trainX,testX,as.factor(trainY), k = 3, l = 0, prob = FALSE, use.all = TRUE)`
- ❖ K-nn with CV - a simpler approach called "leave-out-one" cross-validation can be used, and this is provided by the `knn.cv` function. Using this technique, the observation itself is ignored when looking for its neighbors.
- ❖ SVM
- ❖ NNet

1. K Nearest Neighbor Classification

15000 samples

13 predictor

26 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L', 'M', 'N', 'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W', 'X', 'Y', 'Z'

Pre-processing: centered, scaled

Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 15000, 15000, 15000, 15000, 15000, 15000, ...

Resampling results across tuning parameters:

k	Accuracy	Kappa	Accuracy SD	Kappa SD
3	0.9199277	0.9167123	0.004078469	0.004242216
5	0.9188687	0.9156107	0.003562830	0.003706130
7	0.9178972	0.9146001	0.004523367	0.004705213
9	0.9168892	0.9135518	0.004898951	0.005096635
11	0.9137899	0.9103282	0.004784319	0.004977109

Kappa was used to select the optimal model using the largest value.

The final value used for the model was k = 3.

`confusionMatrix(myknn,testY)`

Overall Statistics

Accuracy : 0.9544

Kappa : 0.9526

2. KNN.CV

```
myknn<-knn.cv(x, y, k = 3, l = 0, prob = FALSE, use.all = TRUE)
```

```
confusionMatrix(myknn, y)
```

Overall Statistics

Accuracy : 0.9614

Kappa : 0.9599

❖ 3. Neural Network

15000 samples

13 predictor

26 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L', 'M', 'N', 'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W', 'X', 'Y', 'Z'

Pre-processing: centered, scaled

Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 15000, 15000, 15000, 15000, 15000, 15000, ...

Resampling results across tuning parameters:

size	decay	Accuracy	Kappa	Accuracy SD	Kappa SD
1	0.0	0.1721552	0.13925465	0.012104369	0.012588550
1	0.1	0.1645228	0.13141653	0.008313172	0.008761456
1	1.0	0.1366458	0.10276125	0.006194946	0.006369236
1	2.0	0.1243592	0.09018002	0.008899931	0.008694901
2	0.0	0.3392758	0.31291087	0.039539183	0.041060387
2	0.1	0.3389576	0.31261560	0.011208979	0.011558657
2	1.0	0.2963639	0.26850521	0.011969575	0.012334720
2	2.0	0.2724686	0.24372387	0.014093756	0.014460833
3	0.0	0.5012276	0.48124893	0.022398048	0.023284106
3	0.1	0.4965047	0.47636823	0.008824991	0.009161699
3	1.0	0.4581531	0.43656710	0.008548176	0.008836588
3	2.0	0.4261495	0.40337037	0.011090971	0.011372202

Kappa was used to select the optimal model using the largest value.

The final values used for the model were size = 3 and decay = 0.

Overall Statistics For Testing set.

Accuracy : 0.4882

Kappa : 0.4675

❖ 4. SVM

15000 samples

13 predictor

26 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L', 'M', 'N', 'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W', 'X', 'Y', 'Z'

Pre-processing: centered, scaled

Resampling: Bootstrapped (25 reps)

Summary of sample sizes: 15000, 15000, 15000, 15000, 15000, 15000, ...

Resampling results across tuning parameters:

C	Accuracy	Kappa	Accuracy SD	Kappa SD
0.25	0.8116716	0.8041039	0.005517341	0.005736462
0.50	0.8429431	0.8366312	0.004796872	0.004988122
1.00	0.8686856	0.8634081	0.004181792	0.004350780
2.00	0.8903813	0.8859756	0.003851433	0.004006828
4.00	0.9079643	0.9042654	0.003491611	0.003631531

Tuning parameter 'sigma' was held constant at a value of 0.02235342

Kappa was used to select the optimal model using the largest value.

The final values used for the model were sigma = 0.02235342 and C = 4.

Overall Statistics for testing set

Accuracy : 0.909

Kappa : 0.9053

Conclusion

Model	Kappa	Accuracy
LDA	0.68	0.67
PLSDA	0.15	0.19
KNN	0.9529	0.9544
KNN-CV	0.9599	0.9614
SVM	0.909	0.905
NeuralNet	0.4882	0.4246