Computer Lab 1 block 1

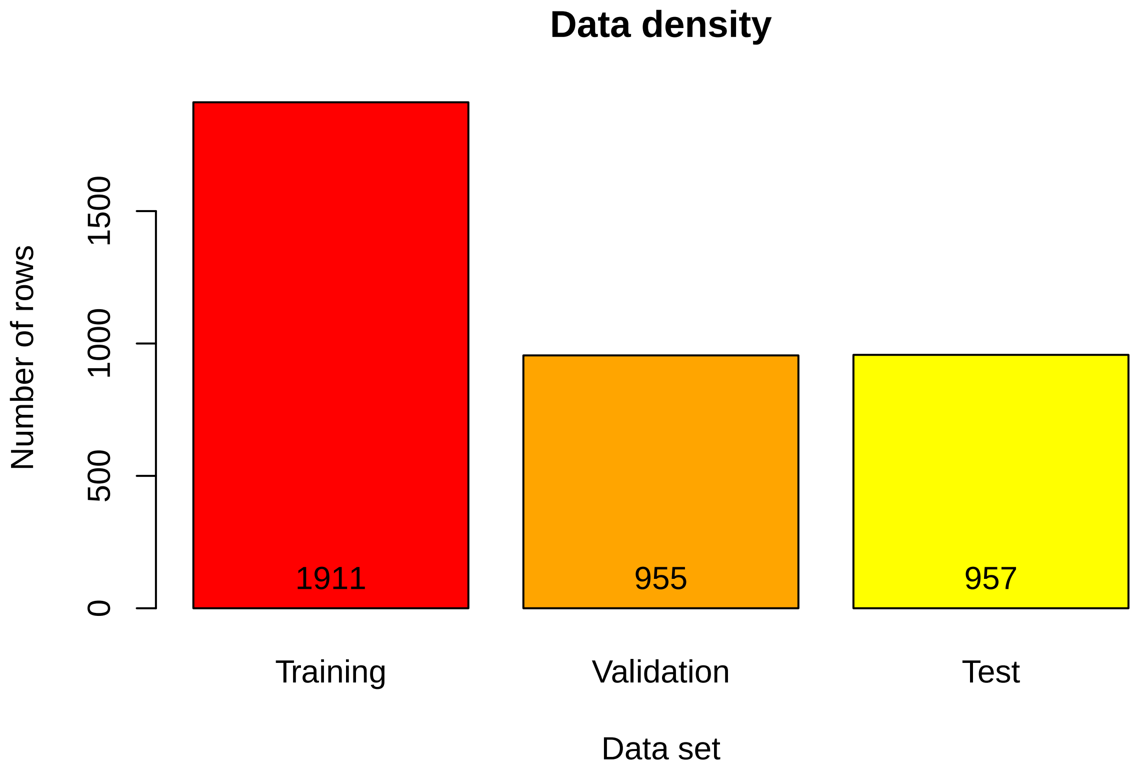
Kerstin, Tugce and Alan

2022-11-16

Assignment 1 Handwritten digit recognition with Knearest neighbors.

1.1 Data set Partition

The original data set **"optdigits.csv"** has a dimension of **3823 rows and 65 columns.** The last column[65] is the target, and represent which number the each row is. After dividing the data set with the proportions indicated the number of rows for each sub set are: **Training: 1911, Validation: 955 and Test: 957**.



1.2 Training the model (Confusion matrices and Misclasification errors)

The Misclassification rate is calculated as follows:

$MR = \frac{FP+FN}{TotalCases}$

- **FP**= False Positives
- **FN** = False Negatives
- **FP + FN**= Incorrect Predictions

The Missclassification Rate of the Training data set is 4.50 % The Missclassification Rate of the Test data set is 5.33 %

From the confusion matrices of Testing and Training data sets, it can be seen that for both models the numbers with most prediction errors are the number 1,7 and 9. From the Testing confusion matrix the number 0 has 100% of accuracy, with no prediction error. In comparison the Training data set has not a 100% of accuracy in a number, but the "best" predicted number is the number 5, followed by numbers 6 and 0.

The Misclasification rate indicates that the model with a better quality of the prediction is the model fitted with the Training data set and test with the same data set. This result is due to the fact that the model is being tested with data that it has already "seen", because it was train with it.

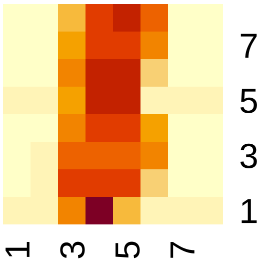
```
##    pred_test
##       0   1   2   3   4   5   6   7   8   9
##   0  77   0   0   0   1   0   0   0   0   0
##   1   0  81   2   0   0   0   0   0   0   3
##   2   0   0  98   0   0   0   0   0   3   0
##   3   0   0   0 107   0   2   0   0   1   1
##   4   0   0   0   0  94   0   2   6   2   5
##   5   0   1   1   0   0  93   2   1   0   5
##   6   0   0   0   0   0   0  90   0   0   0
##   7   0   0   0   1   0   0   0 111   0   0
##   8   0   7   0   1   0   0   0   0  70   0
##   9   0   1   1   1   0   0   0   1   0  85
```

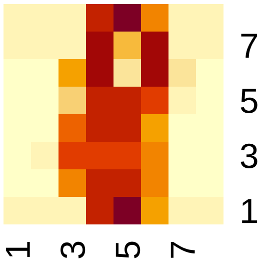
```
##    pred_train
##       0   1   2   3   4   5   6   7   8   9
##   0 202   0   0   0   0   0   0   0   0   0
##   1   0 179  11   0   0   0   0   1   1   3
##   2   0   1 190   0   0   0   0   1   0   0
##   3   0   0   0 185   0   1   0   1   0   1
##   4   1   3   0   0 159   0   0   7   1   4
##   5   0   0   0   1   0 171   0   1   0   8
##   6   0   2   0   0   0   0 190   0   0   0
##   7   0   3   0   0   0   0   0 178   1   0
##   8   0  10   0   2   0   0   2   0 188   2
##   9   1   3   0   5   2   0   0   3   3 183
```

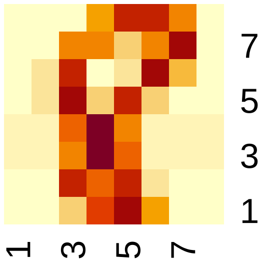
1.3 Easiest and Hardest cases for digit "8"

The heatmap plot shows that the three cases with lowest probability or the hardest to predict, have shapes which the two circles of the 8 figure is not define in comparison with the easiest to predict that both circles are well define. Another detail is that the cases with highest probability have darker colors on the perimeter of the shape, this indicates that there is more pixels in that position and is easier for the model to identify.

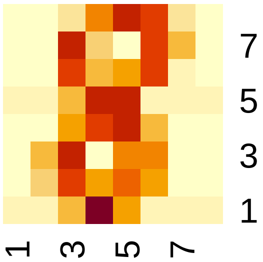
```
##      Index Hardest % Index Easiest %
## [1,]    50 0.1000000   192         1
## [2,]    43 0.1333333   203         1
## [3,]   136 0.1666667    NA        NA
## Hardest
```

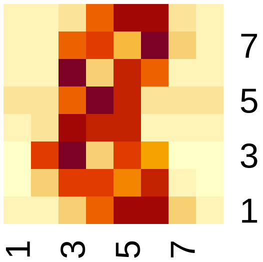






```
## Easiest
```



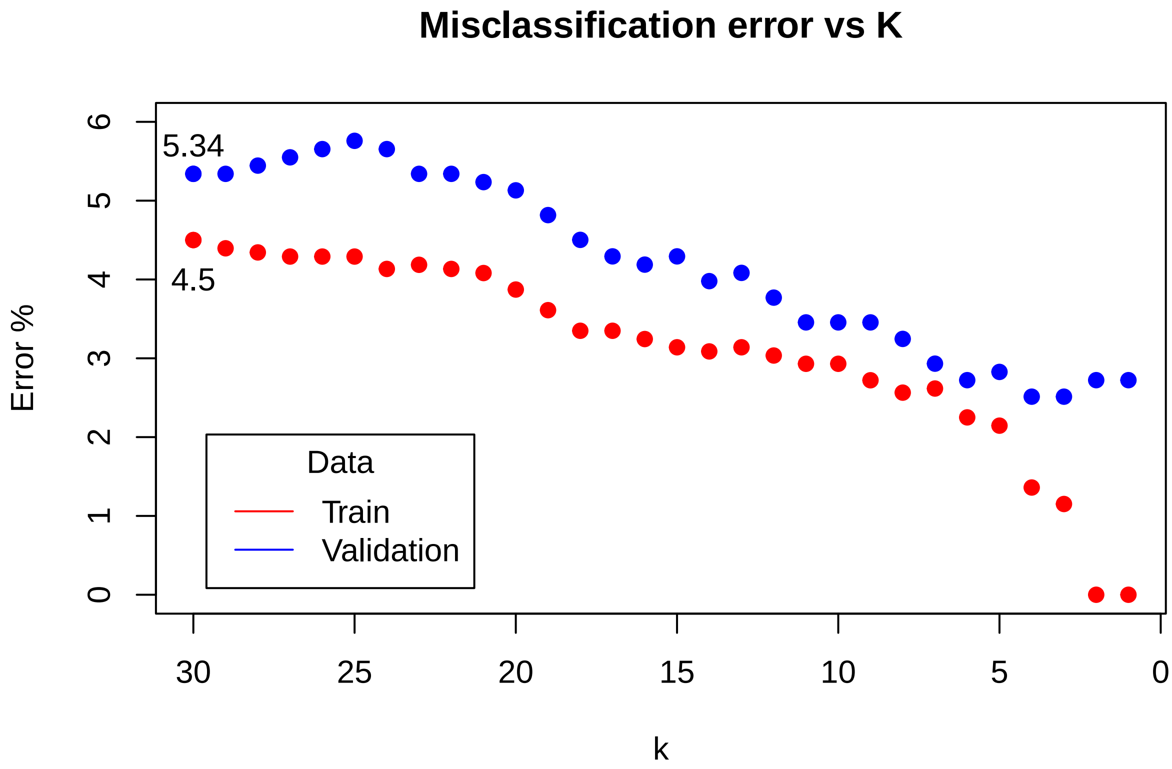


1.4 Model Complexity

After training different models changing the value of k= [1:30], with the training and validation data. From the plot there is a substantial difference between the models, the model train and test with the training data show a smaller Misclassification error than the Validation model. The complexity of the models increases as the n umber of K or numbers nearest neighbors decreases. The optimal K is 4, because it has to be made a trade off between the complexity of the model and the Misclassification error. In this case K equal to 4 is the one with smaller Misclassification rate and without being too small, this means a model with less complexity. We want to choose a model with lower complexity to avoid overfitting.

Each model is different and could have different type of data; for example a model for a health condition is more sensible than one about identifying types of rocks. With that said, it could be choose a K bigger depending of how much impact an small increase on the Misclassification rate creates. So one could choose a different for the case used in this assignment, an optimal K equal to 11, because the error difference against K = 4 is around 1 percent.

From the plot it can be seen that K equals to 1 and 2, for the training data set, resulted in a Misclassification rate equals to 0. In this case with an small number of K, the model will just identified each observation as only point to consider while training, so identifies each observation as single case which could lead to over fitting.

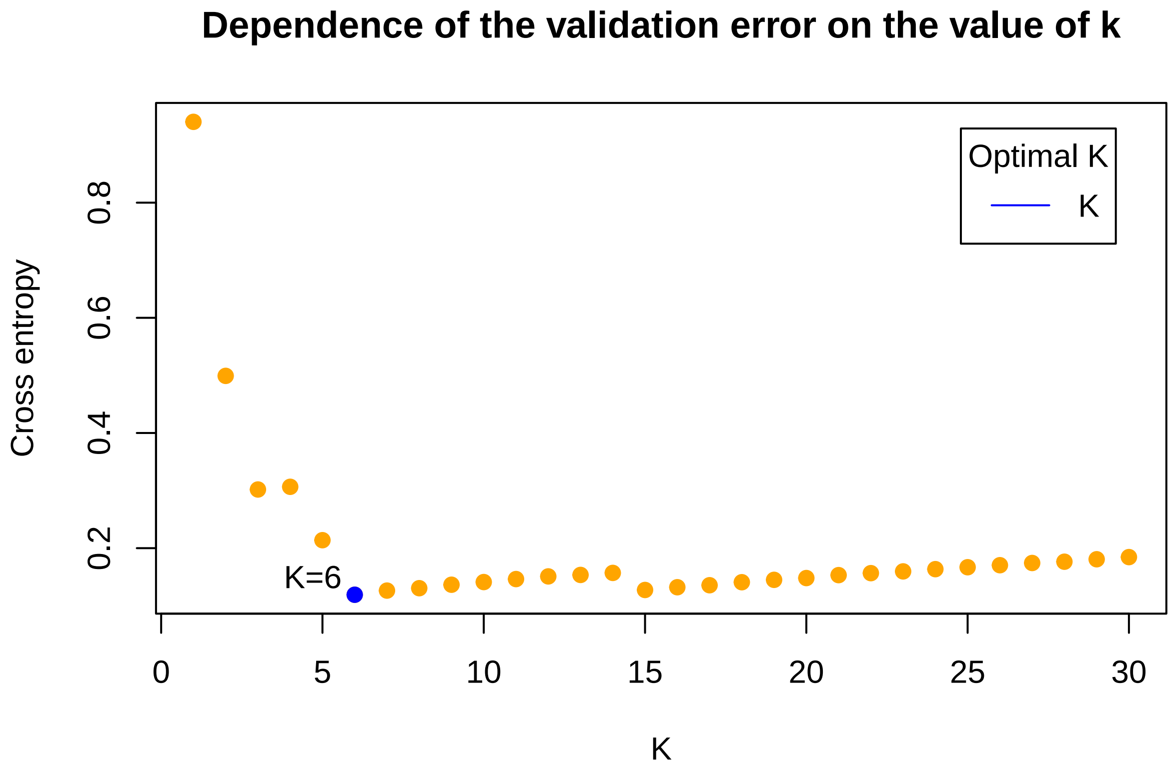


1.4.1 Misclassification error with Test data set, with optimal K The error of the model with the Test data set have a similar value to the validation model, 2.510%(Test) vs 2.513%(Validation). The result is expected because both data sets have a similar amount of data, 955(validation) vs 957(Test).

```
##  Missclasification rate trainning % Missclasification rate validation %
##                               1.360                               2.513
##        Missclassification_rate_Test
##                               2.510
```

5. Cross entropy for training data

The optimal value of K is 6, because is the one with lowest cross entropy and the magnitude of K is not to small to create overfitting. Cross entropy is a better error performance metric than Misinterpretation rate for multiclass classification, because cross entropy deals with the total probability of the targets and gives a better understanding of the whole model. At then end the aim of using the cross entropy is to minimize the distance between the two probability distributions.


Dependence of the validation error on the value of k
Optimal K
K
K=6
Cross entropy
0.2
0.4
0.6
0.8
0
5
10
15
20
25
30
K