Yue Pan
yuepan@student.ethz.ch

Assignment 4

Image Interpretation
November 6, 2019

Exercise 1

Topic: K Nearest Neighbor (KNN) Classification.

Firstly, KNN with different K value, vote weighting rules and distance metrics are tested on a toy dataset.

In this case, parameter setting of K value varying from 1 to 10, the distance metric 'Euclidean (L2 norm)' and 'Cityblock (L1 norm)' as well as the weighting rule 'equal weight', 'distance inverse weight' and 'distance square inverse weight' are tested. Their results are shown from Fig.1 to Fig.4. Note that when there's a tie during the voting process, the point would be assigned to the class with the smallest index.

Since the performance of the algorithm on toy dataset is difficult to evaluate, we can only draw some general idea from the results:

1. The smaller the K value, it is easier to over-fit and is less robust to noise. As shown in the images, we can notice some isolated 'islands' surrounded by the other category's decision area, which is caused by noise.
2. The larger the value of K, it is easier to under-fit. There is no such kind of isolated 'islands' in these cases.
3. Distance metrics and weight rule has little influence on the result of decision boundary.

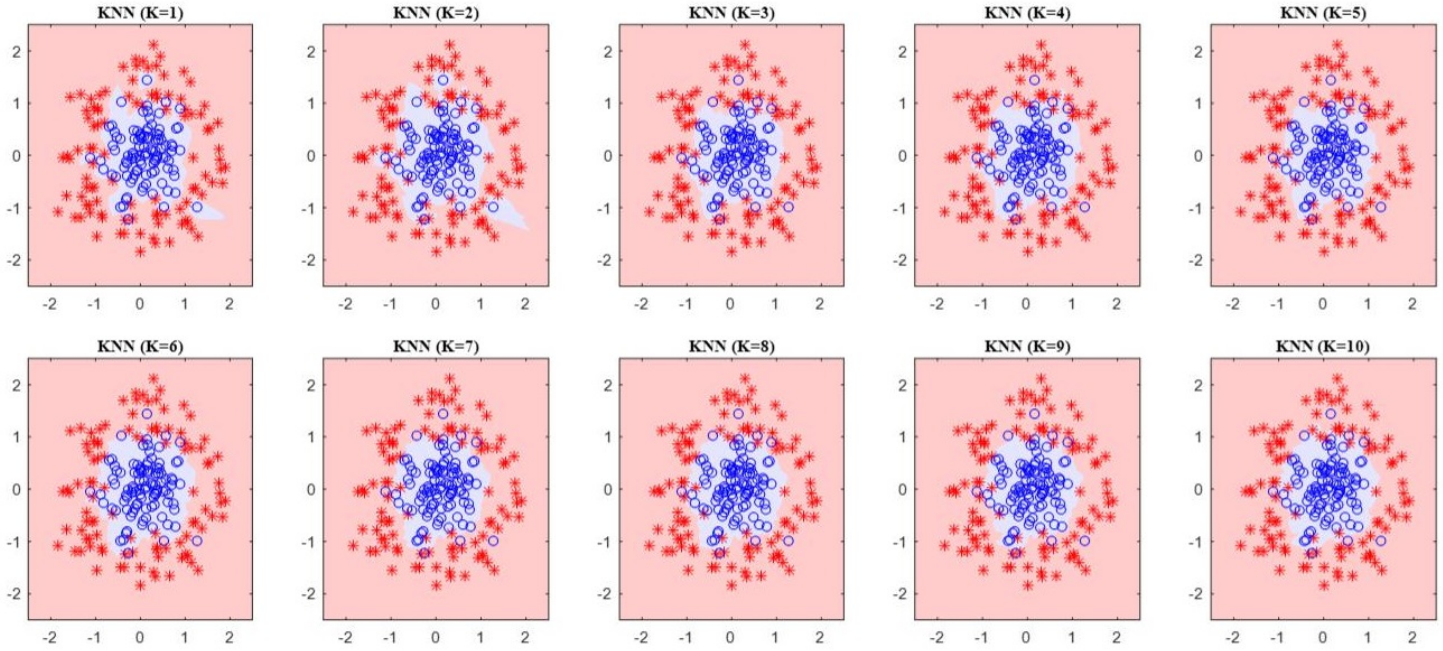

Figure 1: KNN classification on toy dataset w.r.t K (L2 Norm - Euclidean distance metric and equal weight)

Secondly, KNN classification with different parameter settings are tested on the Graz dataset (refer to Fig.5 for Graz dataset description). In this case, the feature space would not be the X and Y coordinate value in toy dataset. Instead, it would be R,G,B value of the image or some other pixel-wise filtered features. Then KNN is implemented in such kind of feature space.

The results for different parameter settings, input features and categories are shown in Fig.6 - Fig.12. The left quarter of the image is used as the training data while the rest pixels are used as testing data. Then for the testing result, sum of difference from the ground truth data, the precision and the recall can be calculated. Precision, recall and F1 score can be calculated according to Eq.1. After comparing the residual, precision and recall with each other, following conclusions are drawn:

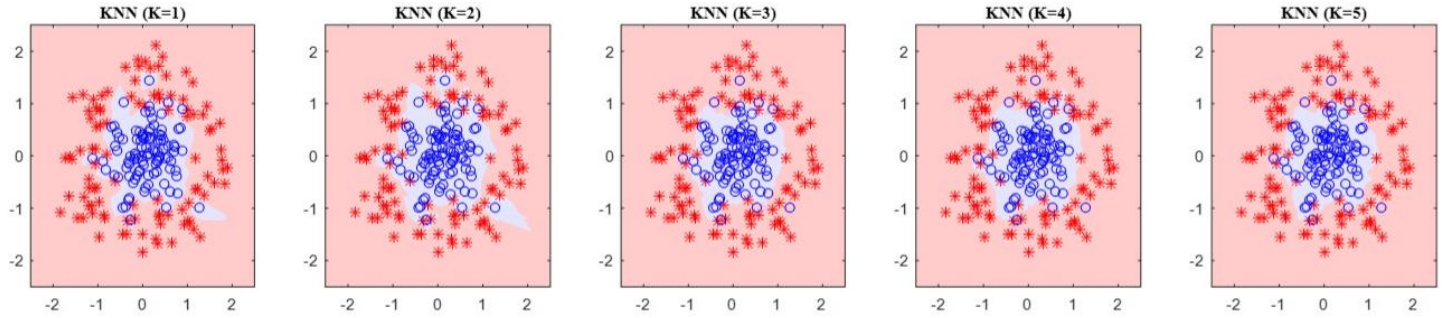

Figure 2: KNN classification on toy dataset w.r.t K (L1 Norm - Manhattan distance metric and equal weight)

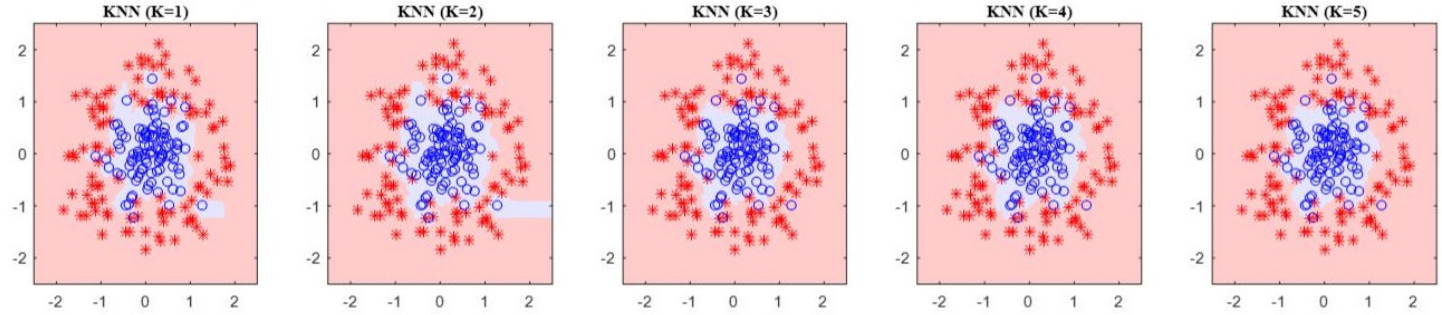

Figure 3: KNN classification on toy dataset w.r.t K (L2 Norm - Euclidean distance metric and inverse distance weight)

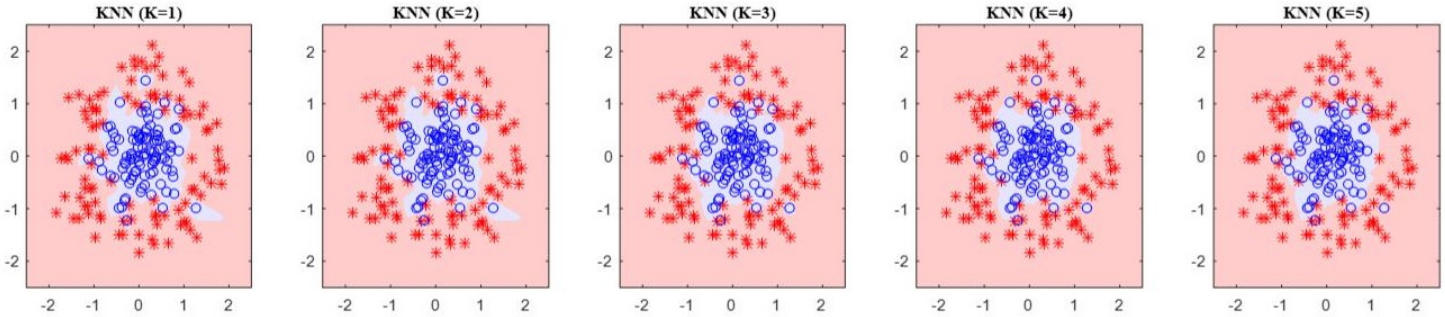

Figure 4: KNN classification on toy dataset w.r.t K (L2 Norm - Euclidean distance metric and inverse distance square weight)

1. When $K < 6$, the precision and recall are relatively low. When $K > 12$, the precision and recall would not increase (and even decrease) while the consuming time increase a lot.

2. Compared with using L2 norm, using L1 norm distance metric has little effect on the results.

3. Compared with using equal weight, using inverse distance or inverse square distance weight makes the result even worse.

4. Compared with using only RGB value as input feature, using also the Gaussian filtered (kernel size set to be 3) feature makes the result even worse.

Figure 5: Graz dataset and its ground truth classification

So after the comparison of classification precision and recall together with consuming time, I get the best parameters setting for KNN:

K=12, Euclidean distance metric, equal weight and only take R,G,B value as input.

$$\begin{cases} \text{precision} = \frac{\text{true positive}}{\text{true positive+false positive}} \\ \text{recall} = \frac{\text{true positive}}{\text{true positive+false negative}} \\ F1 = \frac{2\times\text{precision}\times\text{recall}}{\text{precision+recall}} \end{cases} \tag{1}$$

Exercise 2

Topic: Linear Discriminant Analysis (LDA) Classification.

For this exercise, please refer to *ii_train_lda.m* and *ii_test_lda.m*.

Given two class of training data C_1 and C_2, the training process of LDA can be done through Eq.2 - Eq.4, from which we get the normal θ and threshold t. Then, for an unseen test data j, we can determine its category according to Eq.5.

$$S_w = \sum_{i \in C_1} (x_i - \mu_1)(x_i - \mu_1)^T + \sum_{i \in C_2} (x_i - \mu_2)(x_i - \mu_2)^T \tag{2}$$

$$\underset{K\times 1}{\theta} = \underset{K\times K}{S_w}^{-1} \underset{K\times 1}{(\mu_2 - \mu_1)} \tag{3}$$

$$\underset{1\times 1}{t} = \underset{1\times K}{\left(\frac{1}{2}(\mu_2 + \mu_1)\right)^T} \underset{K\times 1}{\theta} \tag{4}$$

$$j \in \begin{cases} C_1 & , x_j^T\theta < t \\ C_2 & , x_j^T\theta > t \end{cases} \tag{5}$$

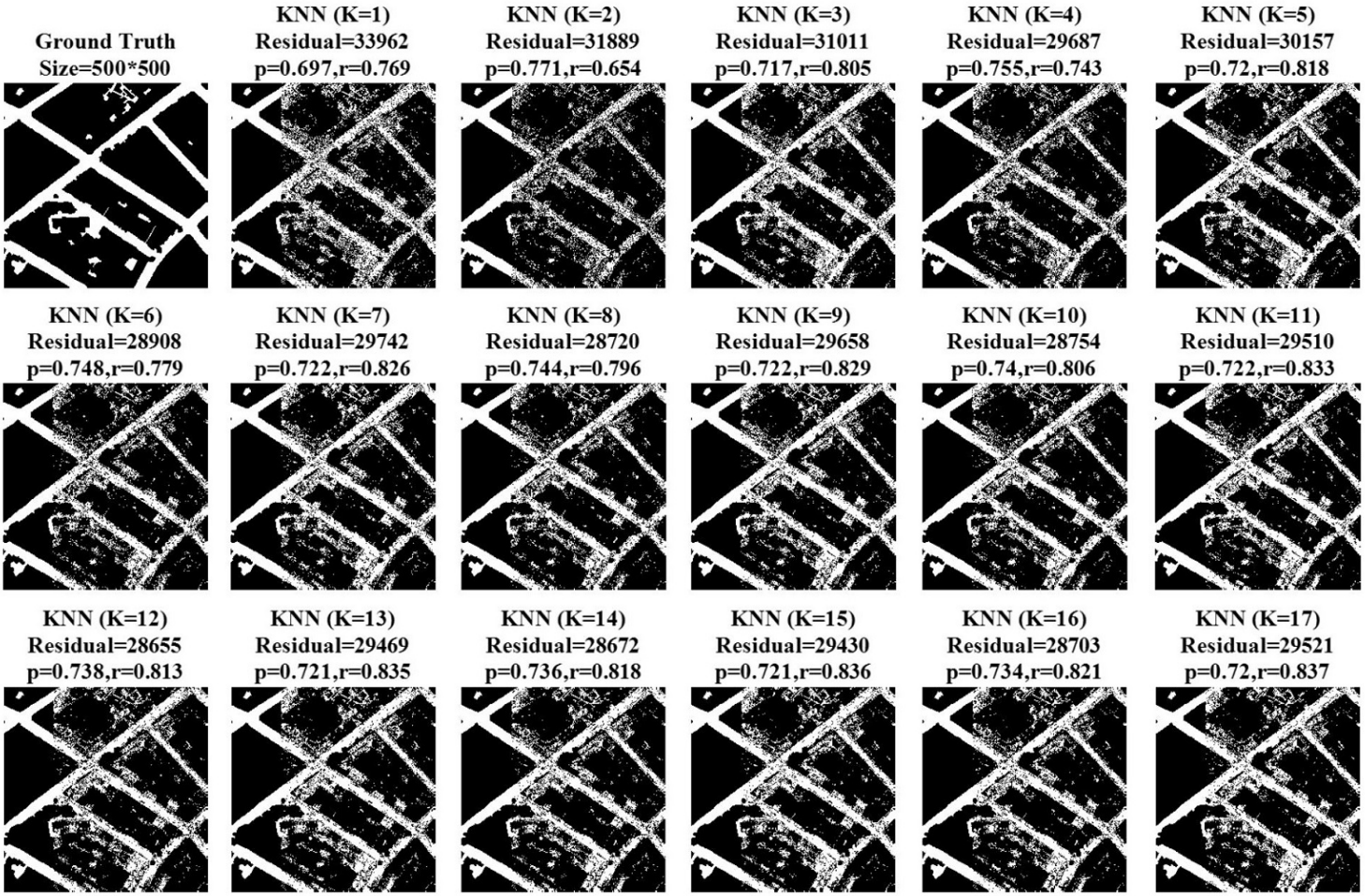

Figure 6: Graz dataset road classification result using KNN w.r.t K (L2 Norm - Euclidean distance metric and equal weight with RGB value as input feature)

For the toy dataset, the non-lifted and lifted results are shown in Fig.13. For the Graz dataset, different features (R,G,B and Gaussian) are set as the input. For different Gaussian kernel size (the standard deviation is set as 1/6 of the kernel size according to Assignment 2), different classification result would be generated, as shown in Fig.14 - Fig.16. It is found that when the gaussian kernel's size $s = 3$, best performance would be achieved, which indicates that more features than just R,G,B value can improve the classification to a certain extent for LDA.

Exercise 3

Topic: Logistic Regression Classification.

For this exercise, please refer to *ii_train_logistic.m*, *ii_test_logistic.m* and *classify_breast_cancer_with_logress*.

In matlab, *glmval* and *glmfit* act as the training and testing function of logistic regression. 'Binomial' distribution and 'logit' 'link' function are set in *glmfit*.

For the toy dataset, the non-lifted and lifted results are shown in Fig.17. For the Graz dataset, different

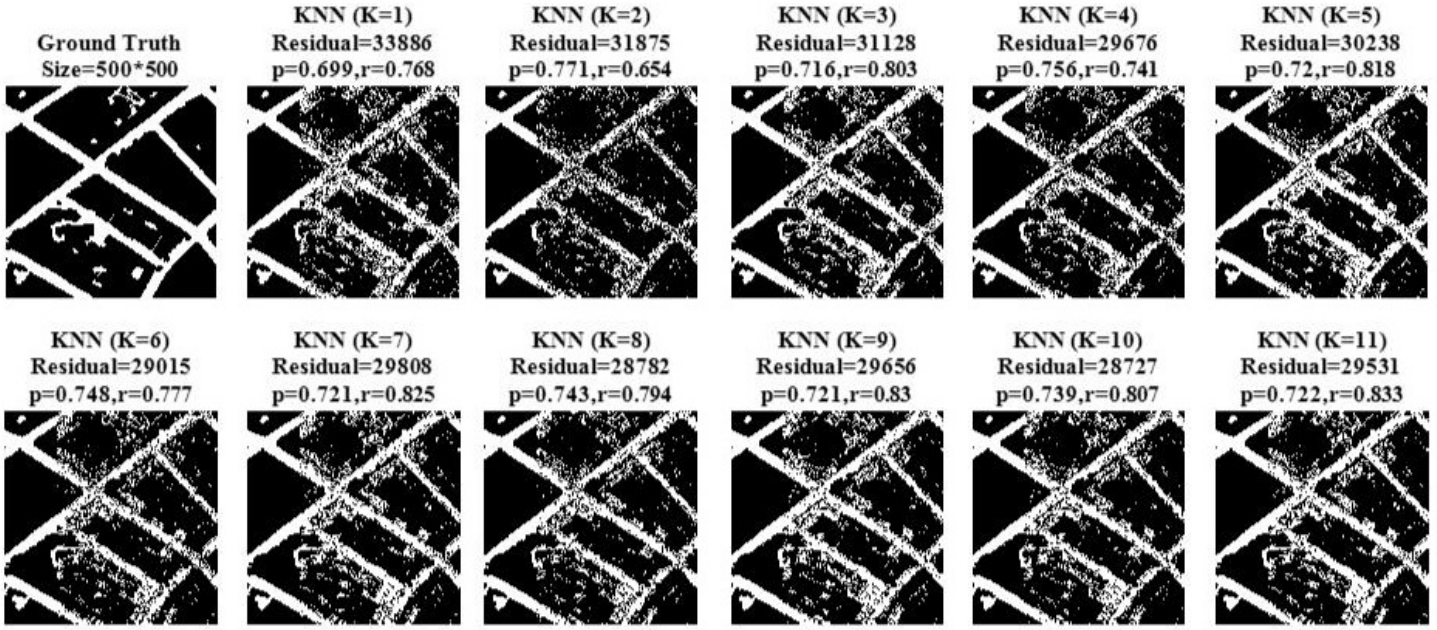

Figure 7: Graz dataset road classification result using KNN w.r.t K ((L1 Norm - Manhattan distance metric and equal weight with RGB value as input feature)

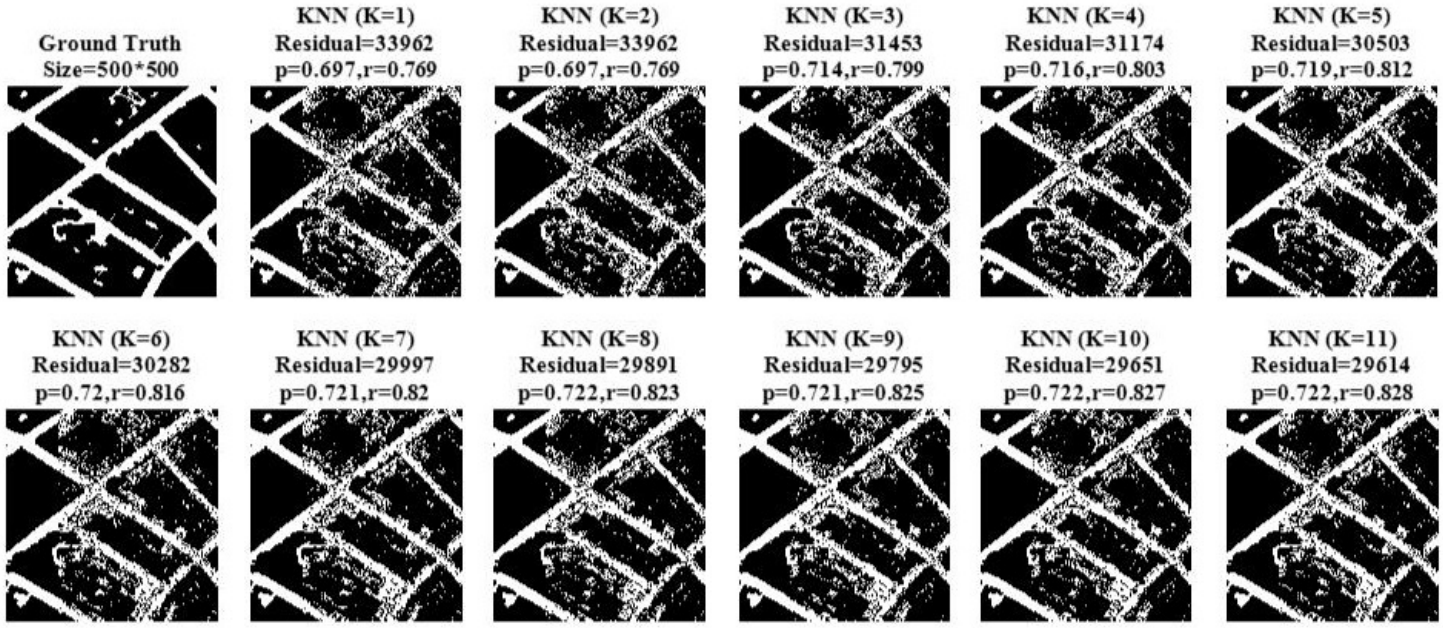

Figure 8: Graz dataset road classification result using KNN w.r.t K (L2 Norm - Euclidean distance metric and inverse distance weight with RGB value as input feature)

features (R,G,B and Gaussian) are set as the input. For different Gaussian kernel size (the standard deviation is set as 1/6 of the kernel size according to Assignment 2), different classification result would be generated for each category, as shown in Fig.18 - Fig.20. It is found that when the gaussian kernel's size $s = 7$, best

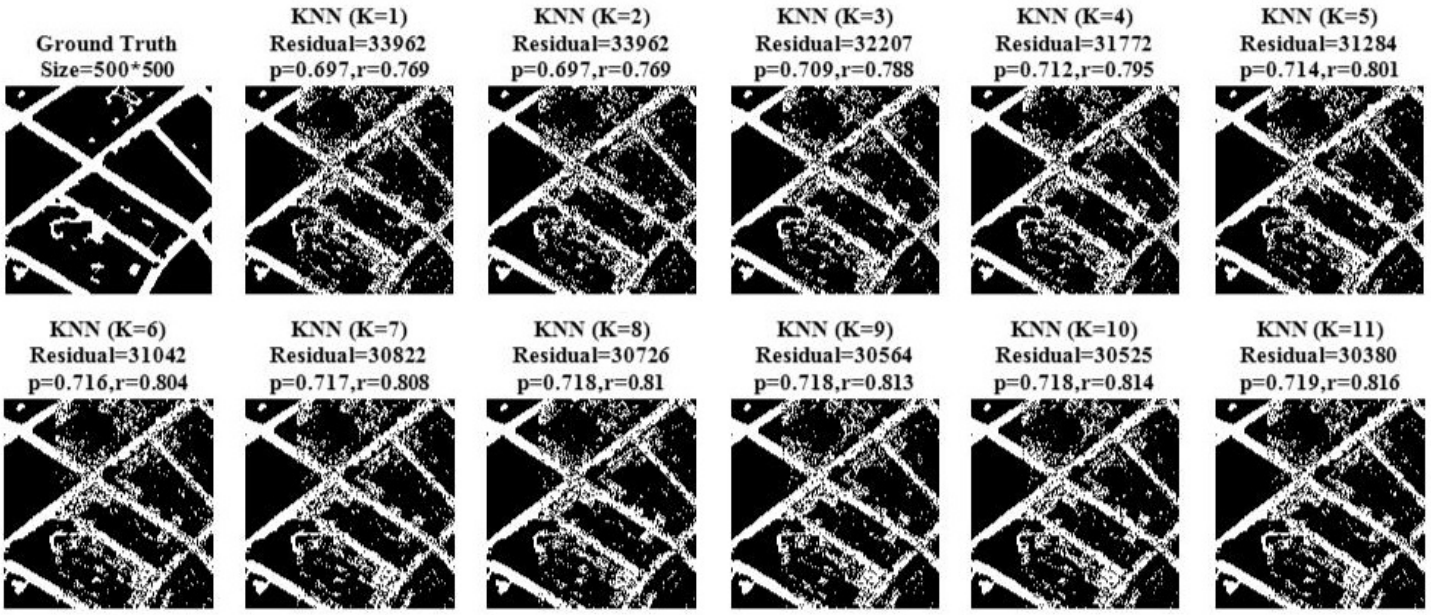

Figure 9: Graz dataset road classification result using KNN w.r.t K (L2 Norm - Euclidean distance metric and inverse distance square weight with RGB value as input feature)

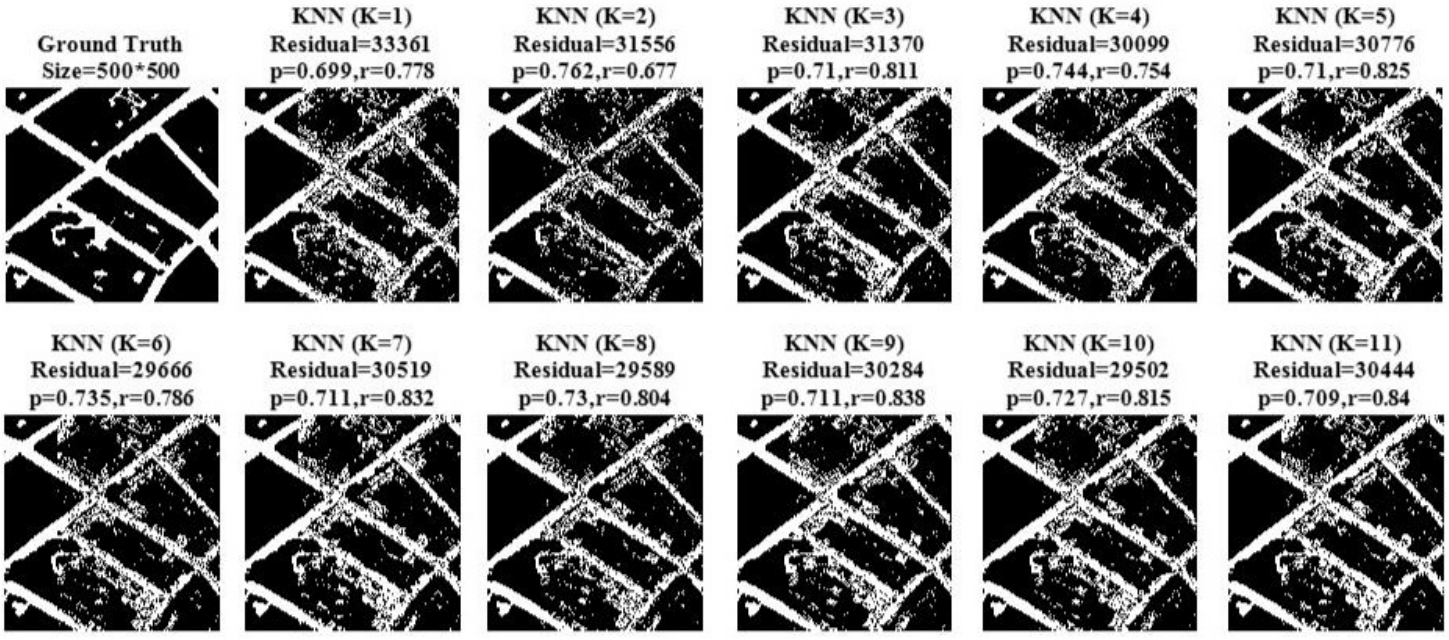

Figure 10: Graz dataset road classification result using KNN w.r.t K (L2 Norm - Euclidean distance metric and equal weight with RGB value and Gaussian filter result as input feature)

performance would be achieved, which indicates that more features than just R,G,B value can improve the classification to a certain extent for Logistic regression.

For the breast cancer dataset, it is found that there's no missing data in the table so that no preprocessing

Figure 11: Graz dataset building classification result using KNN w.r.t K (L2 Norm - Euclidean distance metric and equal weight with RGB value as input feature)

needs to be done. For better evaluate the classification performance, a 4-folds (75% training data) cross validation is adopted here. The mean accuracy for 4 testing sets is 0.9662.

Exercise 4

Topic: Decision Tree Classification.

For this exercise, please refer to *ii_train_tree.m*, *ii_test_tree.m* and *classify_breast_cancer_with_tree*.

In matlab, the *fitctree* and *predict* functions are used as the training and testing function for decision tree. Other parameters are set as default value.

For the toy dataset, the non-lifted and lifted results are shown in Fig.21. The corresponding decision tree is shown as Fig.22. It can be noticed that decision tree's decision boundary is much more discrete and regular compared with LDA and logistic regression.

For the Graz dataset, different features (R,G,B and Gaussian) are set as the input. For different Gaussian kernel size (the standard deviation is set as 1/6 of the kernel size according to Assignment 2), different classification result would be generated for each category, as shown in Fig.23 - Fig.25. It is found that when

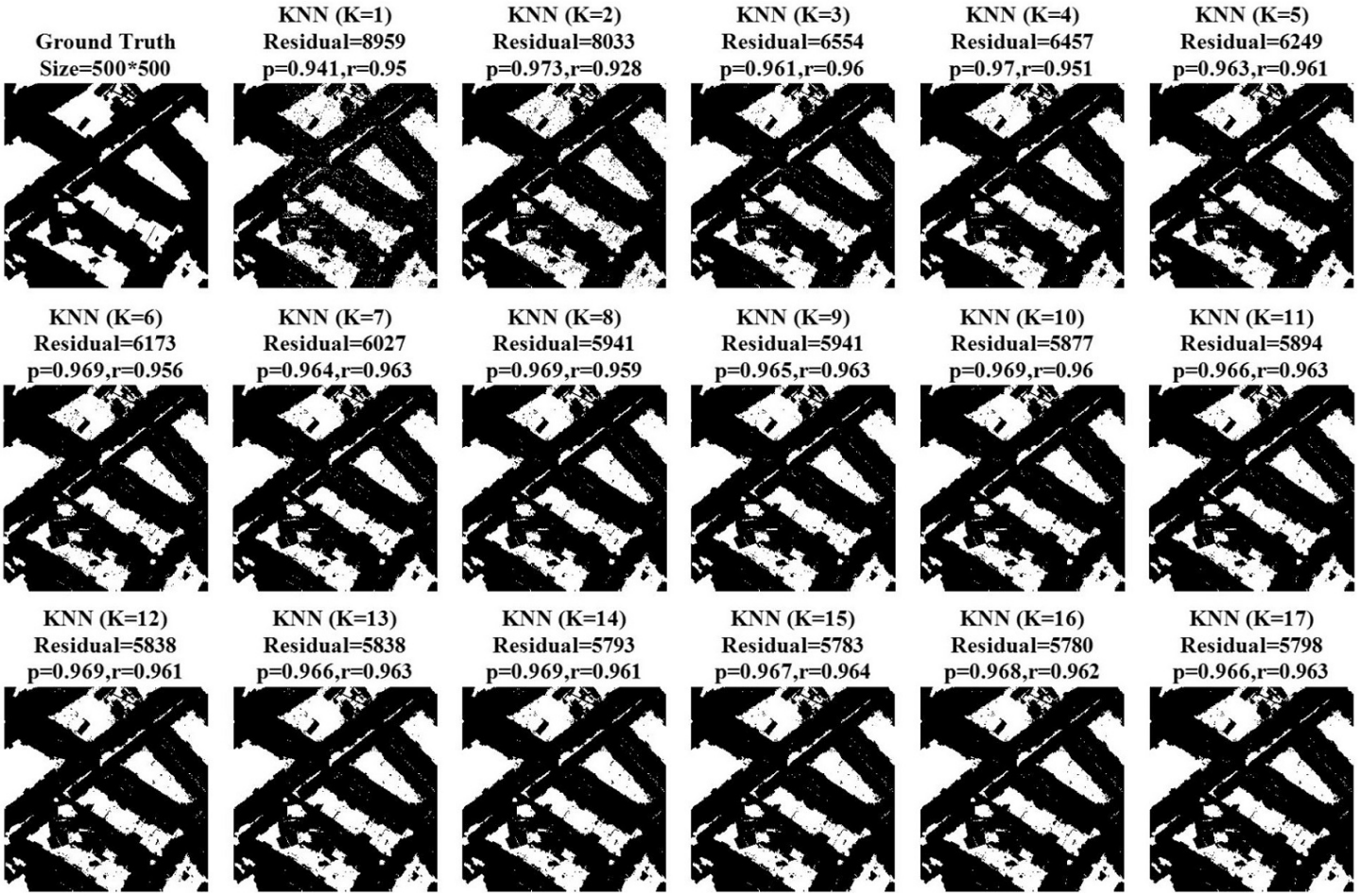

Figure 12: Graz dataset vegetation classification result using KNN w.r.t K (L2 Norm - Euclidean distance metric and equal weight with RGB value as input feature)

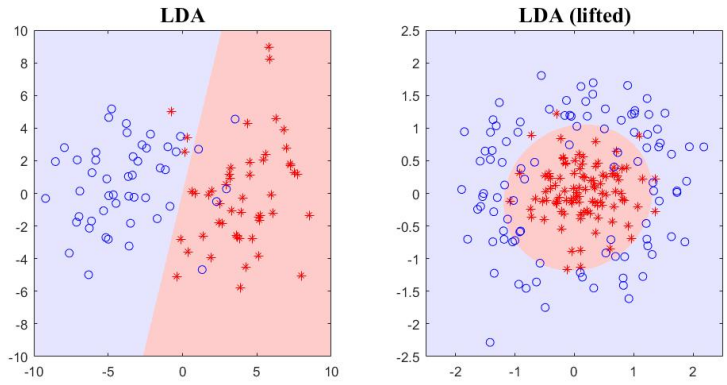

Figure 13: LDA Classification on toy dataset

only R,G,B value are used as input features, best performance would be achieved, which indicates that more features than just R,G,B value may not improve the classification for decision tree.

All the 4 methods (KNN, LDA, Logistic regression and decision tree) are tested on Graz dataset so that

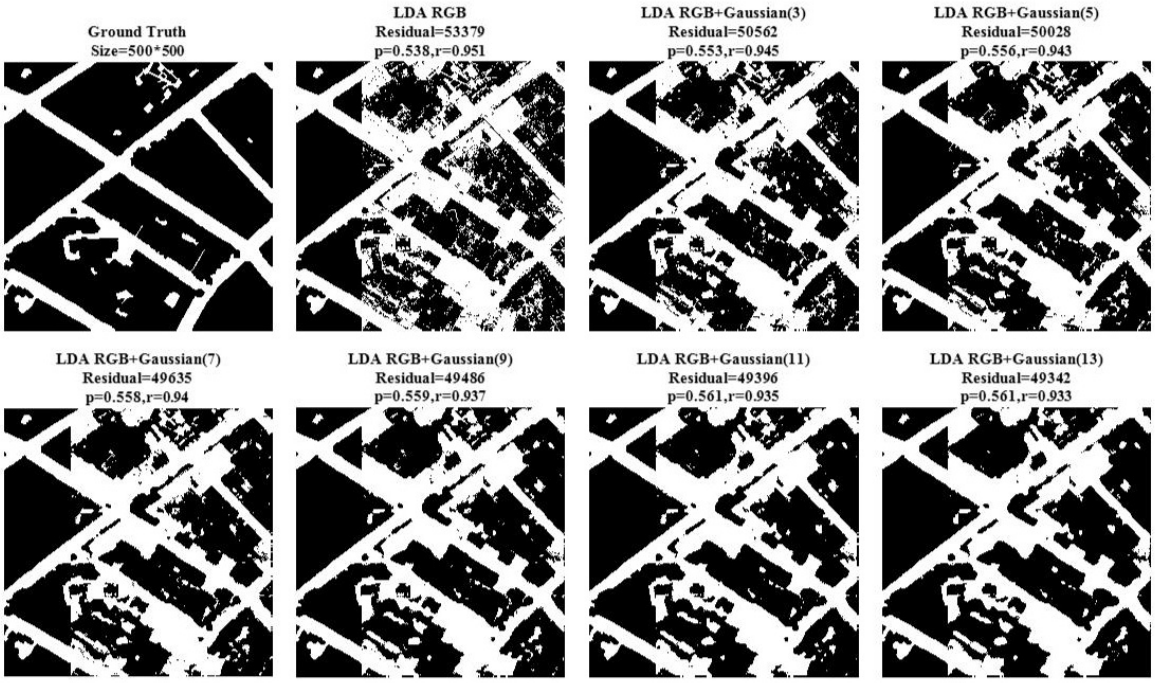

Figure 14: Graz dataset buildings classification result using LDA with different input features

Figure 15: Graz dataset roads classification result using LDA with different input features

we can compare their performance. For each method, its best parameter setting on Graz dataset is used. The comparison results are reported in Table 1. It is found that for Graz Dataset, the most simple method KNN (K=12, euclidean distance metric and equal weight) has the best performance, followed by decision tree.

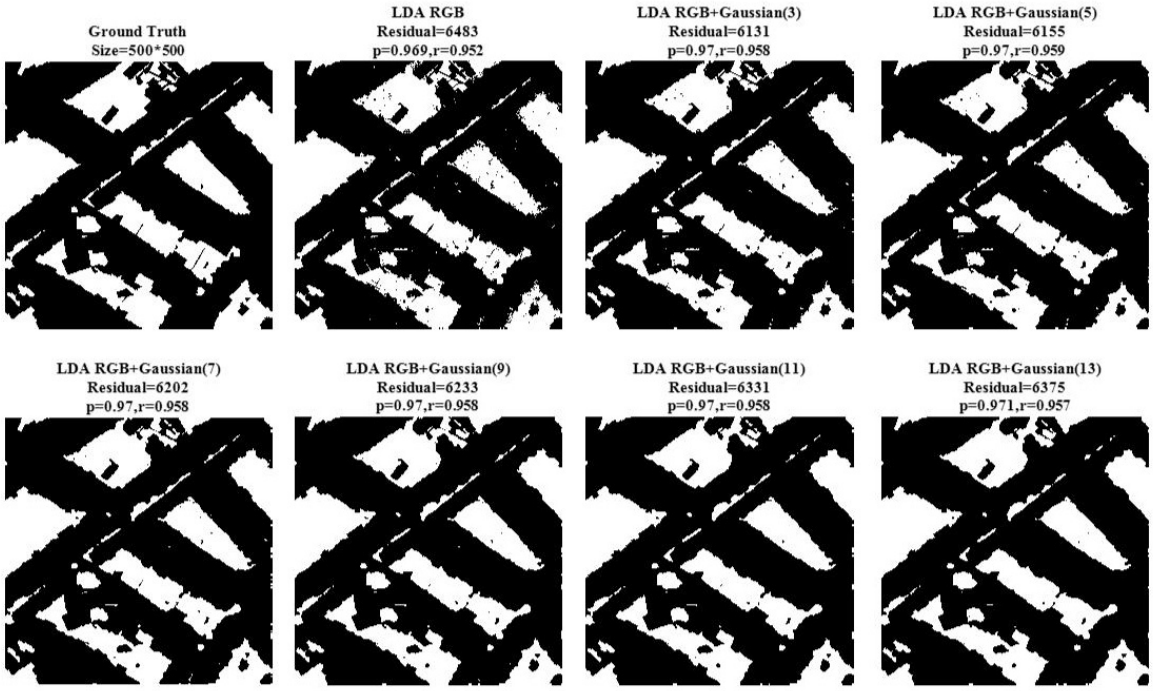

Figure 16: Graz dataset vegetation classification result using LDA with different input features

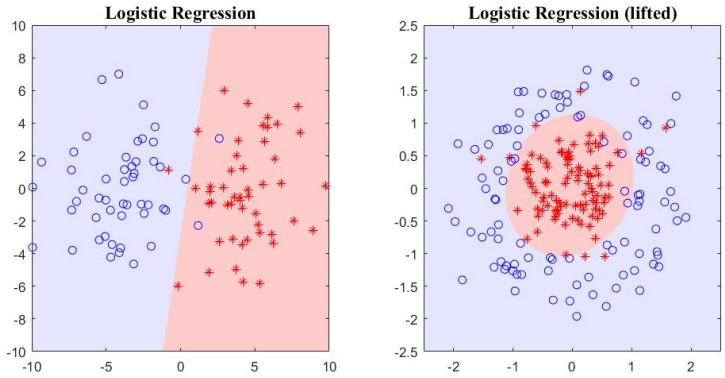

Figure 17: Logistic regression classification on toy dataset

Specifically, for road and building classification, KNN is still the best. However, for vegetation classification, Logistic regression has the best performance.

For the breast cancer dataset, a 4-folds (75% training data) cross validation is adopted. Decision tree's mean accuracy on 4 testing sets is 0.9412, which is a bit lower than Logistic regression's. Logistic regression and decision tree's accuracy are reported in Table 2.

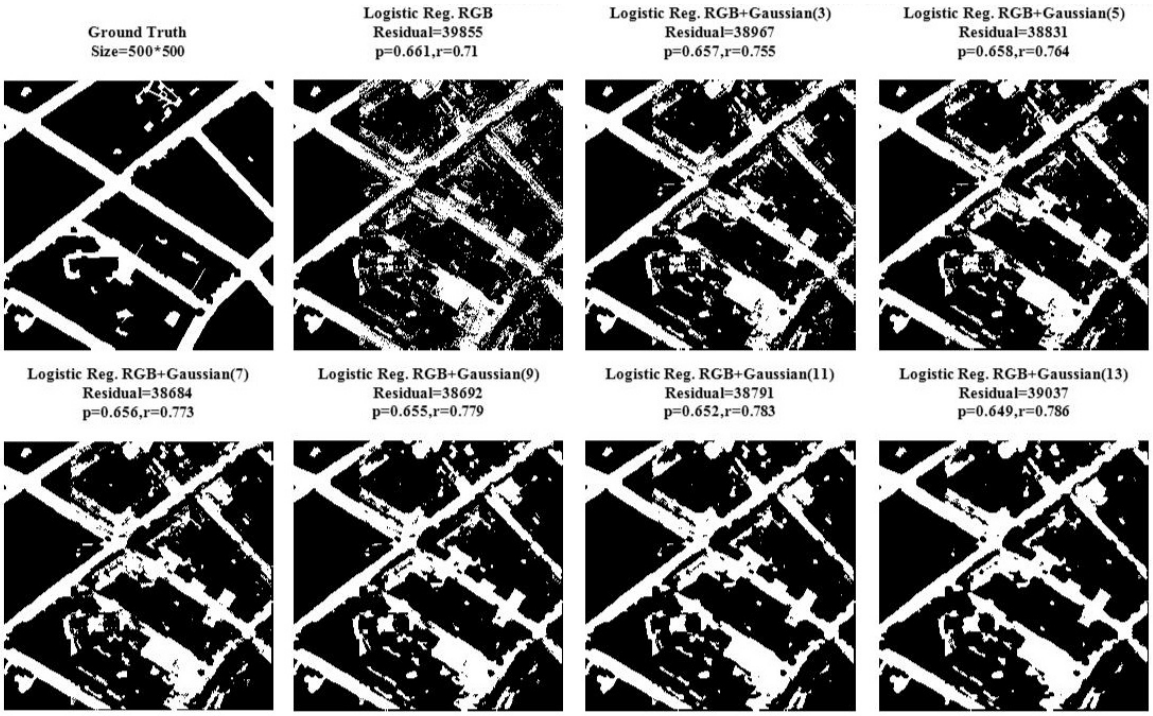

Figure 18: Graz dataset buildings classification result using Logistic Regression with different input features

Figure 19: Graz dataset roads classification result using Logistic Regression with different input features

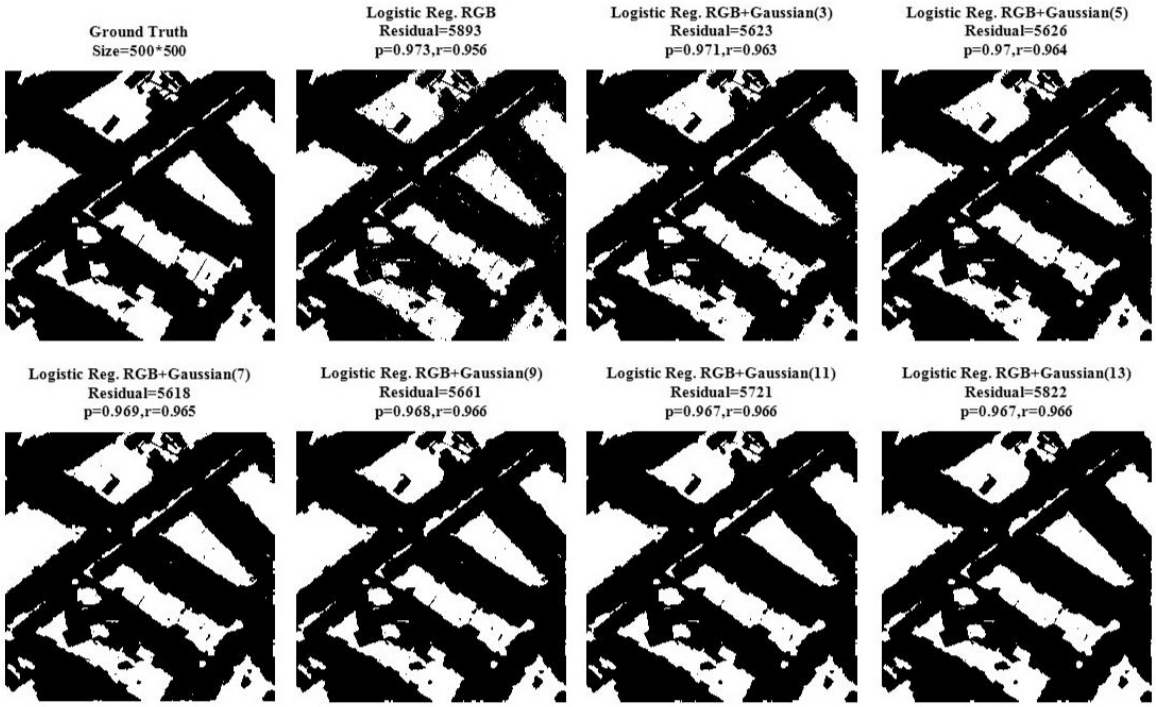

Figure 20: Graz dataset vegetation classification result using Logistic Regression with different input features

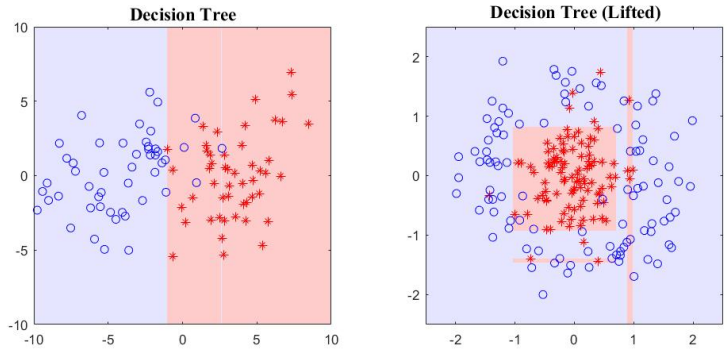

Figure 21: Decision tree classification on toy dataset

Table 1: Classification accuracy evaluation on Graz dataset for all 4 classification methods

Method	Parameters	road			building			vegetation			mean
		p	r	f1	p	r	f1	p	r	f1	m f1
KNN	K=12, L2	**0.738**	0.813	**0.774**	0.924	0.793	**0.854**	0.969	0.961	0.965	**0.864**
LDA	Gau. s=3	0.553	**0.945**	0.698	**0.932**	0.785	0.852	**0.970**	0.958	0.964	0.838
Log.R.	Gau. s=7	0.656	0.773	0.710	0.925	0.771	0.841	0.969	**0.965**	**0.967**	0.839
Tree	None	0.704	0.772	0.736	0.869	**0.812**	0.836	0.945	0.949	0.947	0.841

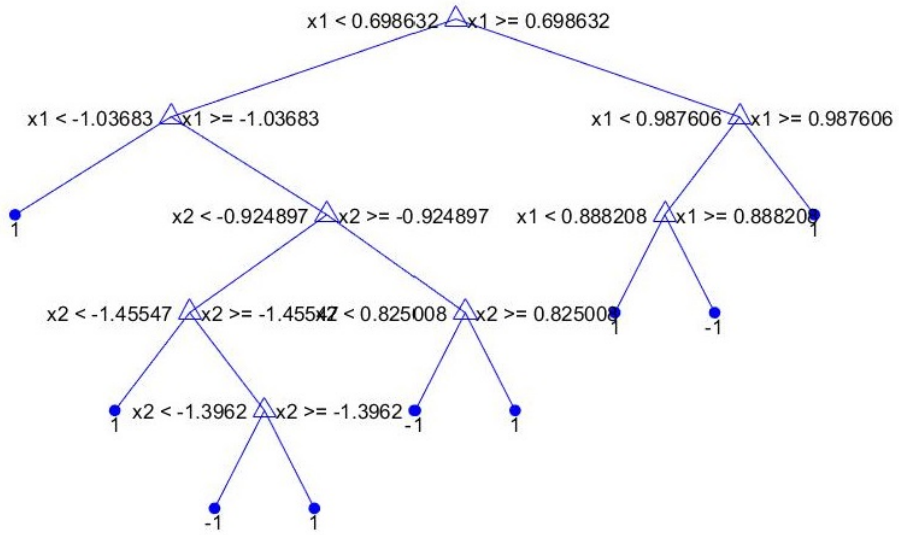

Figure 22: Generated decision tree example (corresponding to Fig.21's lifted case)

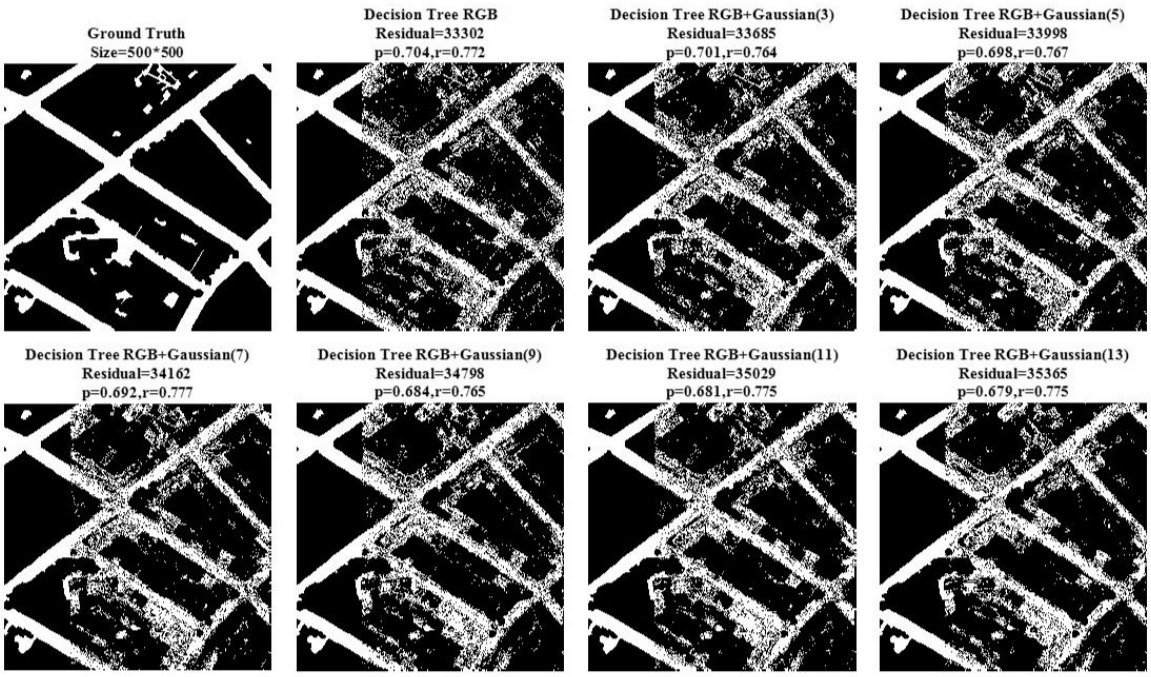

Figure 23: Graz dataset roads classification result using Decision Tree with different input features

Table 2: Classification accuracy evaluation on breast cancer dataset for logistic regression and decision tree using 4-folds cross validation

Method	Logistic regression	Decision tree
Accuracy	**0.9662**	0.9412

Figure 24: Graz dataset buildings classification result using Decision Tree with different input features

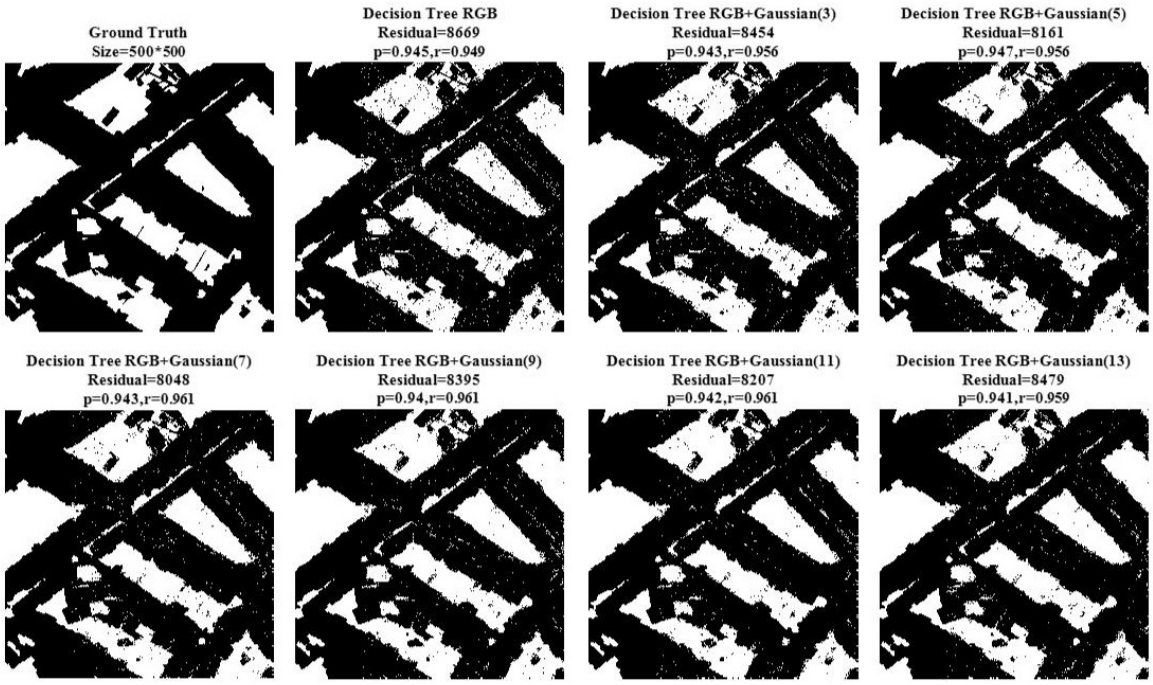

Figure 25: Graz dataset vegetation classification result using Decision Tree with different input features