Exercise 6: Tree variable imputation

1) Read “TreeVariableImputation.csv” into R

2) Predict mean diameter (d1.3), height and tree species for each tree using the ALS derived features.

Use k-NN method (R-package yaImpute). Get familiar with the following document Crookston

& Finley (2008) which is in Moodle.

• Test different explanatory variables (x)

• Test different values of k

• Test different strategies to select the nearest neighbours:

o Euclidean distance

o Most similar neighbour (MSN)

3) Analyze the results

• Calculate RMSE and bias

• Plot dependencies between measured and estimated attributes

Return pdf-document which includes your analysis of the achieved results. Analyze, what kind of dependencies were visible between the ALS derived features and field measured tree-level attributes. What were the best explanatory variables (ALS features) for mean diameter and height?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NAMING | K | Method for searching kNN | Impute K | Impute method | Impute method factor |
| euc1 |  | euclidean |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

# Calculate RMSE

euc1:

k=3, method = "euclidean" (method of searching the k)

k=3, method="dstWeighted", method.factor="median"

euc2<-rmsd.yai(euck5\_Imp, scale=F)

k=5, method = "euclidean" (method of searching the k)

, k=5, method="dstWeighted", method.factor="median")

euc3<-rmsd.yai(euc\_Imp2, scale=F)

k=5, method = "euclidean" (method of searching the k)

euc, k=3, method="mean", method.factor="closest")

euc4<-rmsd.yai(euck5\_Imp2, scale=F)

k=5, method = "euclidean")

k=5, method="mean", method.factor="mean")

msn1<-rmsd.yai(msn\_Imp, scale=F)

(x = x, y = y, k=3, method = "msn")

, k=3, method="dstWeighted", method.factor="median")

msn2<-rmsd.yai(msnk5\_Imp, scale=F)

(x = x, y = y, k=5, method = "msn")

k=5, method="dstWeighted", method.factor="median")

msn3<-rmsd.yai(msn\_Imp2, scale=F)

(x = x, y = y, k=3, method = "msn")

(msn, k=3, method="mean", method.factor="closest")

msn4<-rmsd.yai(msnk5\_Imp2, scale=F)

(x = x, y = y, k=5, method = "msn")

msnk5\_Imp2 <- impute(msnk5, k=5, method="mean", method.factor="mean")

rf<-rmsd.yai(rf\_Imp, scale=F)

yai(x = x, y = y2, k=3, method = "randomForest")

(rf, k=3, method="mean", method.factor="closest")

BIAS DIAMETER

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | euc1 | euc2 | euc3 | euc4 | msn1 | msn2 | msn3 | msn4 | rf |
| diameter bias | 5.201636 | 3.602812 | 5.342949 | 0.225918 | 1.184429 | 1.184429 | 0.195513 | 1.344231 | 3.301282 |

BIAS HEIGHT

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | euc1 | euc2 | euc3 | euc4 | msn1 | msn2 | msn3 | msn4 | rf |
| height bias | 0.30105 | 0.223005 | 0.312527 | 0.014044 | 0.053073 | 0.053073 | 0.01583 | 0.063252 | 0.171088 |

CORRELATION

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | euc1 | euc2 | euc3 | euc4 | msn1 | msn2 | msn3 | msn4 | rf |
| d13\_2009 | 0.700007 | 0.714972 | 0.700615 | 0.709865 | 0.72576 | 0.742723 | 0.725377 | 0.739163 | 0.751047 |
| h | 0.752019 | 0.779727 | 0.756283 | 0.779215 | 0.783561 | 0.802102 | 0.784839 | 0.800756 | 0.794997 |
| Species2 | NA | NA | NA | NA | 0.433937 | 0.522126 | 0.436149 | 0.527491 | NA |

RMSE

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | euc1 | euc2 | euc3 | euc4 | msn1 | msn2 | msn3 | msn4 | rf |
| d13\_2009 | 37.15376 | 36.15569 | 37.10373 | 36.46483 | 35.92286 | 34.47041 | 35.8647 | 34.66948 | 34.27145 |
| h | 1.964529 | 1.866156 | 1.9527 | 1.87528 | 1.848551 | 1.759918 | 1.839827 | 1.767138 | 1.798952 |
| Species2 | NA | NA | NA | NA | 0.474584 | 0.433132 | 0.473665 | 0.43101 | NA |