Predict the type of forest cover from cartographic variables

Variables:

Quantitative

1. Elevation /meters / Elevation in meters
2. Aspect / azimuth / Aspect in degrees azimuth
3. Slope / degrees / Slope in degrees
4. Horizontal\_Distance\_To\_Hydrology / meters / Horz Dist to nearest surface water features
5. Vertical\_Distance\_To\_Hydrology / meters / Vert Dist to nearest surface water features
6. Horizontal\_Distance\_To\_Roadways / meters / Horz Dist to nearest roadway
7. Hillshade\_9am / 0 to 255 index / Hillshade index at 9am, summer solstice
8. Hillshade\_Noon / 0 to 255 index / Hillshade index at noon, summer soltice
9. Hillshade\_3pm / 0 to 255 index / Hillshade index at 3pm, summer solstice
10. Horizontal\_Distance\_To\_Fire\_Points // meters / Horz Dist to nearest wildfire ignition points

Qualitative

1. Wilderness\_Area (4 binary columns) / qualitative / 0 (absence) or 1 (presence) / Wilderness area designation
2. Soil\_Type (40 binary columns) / qualitative / 0 (absence) or 1 (presence) / Soil Type designation

Integer (To be predicted)

1. Cover\_Type (7 types) / integer / 1 to 7 / Forest Cover Type designation 30 x 30 meter cell

* Spruce/Fir (1)
* Lodgepole Pine (2)
* Ponderosa Pine (3)
* Cottonwood/Willow (4)
* Aspen (5)
* Douglas-fir (6),
* Krummholz (7)

Data:

Subset of:

http://archive.ics.uci.edu/ml/datasets/Covertype

Training data: 50000

Test data: 100000

Methods

Tree based methods:

* Random forest
* Gradient boosting

Other:

* SVM
* KNN

Next step:

* Sync github
* Data viz
* Features ?

Other information:

Article:

* The results of the comparison indicated that a feedforward artificial neural network (ann) model more accurately predicted forest cover type than did a traditional statistical model based on Gaussian discriminant analysis.

Features:

* Distance to nearest surface water = feature sqrt(horizontal^2 +vertical ^2)
* Combined variable with slope and sunlight (possibly already included in the sunlight variable)

Tested for predictive value in article and adopted or dropped in the analysis:

* (horizontal distance to the nearest wildfire ignition point
* wilderness area designation
* Soil type designation
* Creating a variable oft he forrest type located next to you

Classification matrix

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Observed / Predicted | Spruce/Fir (1) | Lodgepole Pine (2) | Ponderosa Pine | Cottonwood/Willow | Aspen | Douglas-fir | Krummholz |
| Spruce/Fir (1) |  |  |  |  |  |  |  |
| Lodgepole Pine (2) |  |  |  |  |  |  |  |
| Ponderosa Pine |  |  |  |  |  |  |  |
| Cottonwood/Willow |  |  |  |  |  |  |  |
| Aspen |  |  |  |  |  |  |  |
| Douglas-fir |  |  |  |  |  |  |  |
| Krummholz |  |  |  |  |  |  |  |