Project Forestcover

Team Members

Task:

Predicting forest cover type from cartographic variables. There are 12 features and 7 categories of forest covers. We use randomly picked 75% of the data samples as training data set and the rest 25% for testing. We use three different classification algorithms: aritificial neural networks, ...

Method 1: ANN

§ Plain ANN

- Structure of the ANN: We choose to have 3 layers(1 hidden layer), 5 hidden nodes, and 7 output nodes, and use L2 regularization.

- Choice of the Activation Functions We use both sigmoid function
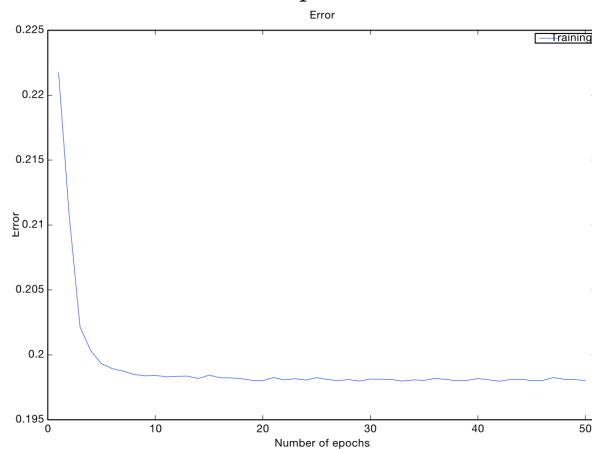
$$S(t) = \frac{1}{1 + e^{-t}} \tag{1}$$

  and tanh function

$$T(t) = 1.7159 tanh(\frac{2}{3}t) \tag{2}$$

  as our activation function.

- Predicting Results Using 12 features

  - Sigmoid function: Train: 0.198, Test: 0.276
  - Tanh function: Train: 0.196, Test: 0.27

- Here is an Error Rate plot

- Here is a ROC plot

- 7 AUC and weighted AUC

§ Alternative Structure ANN

For ANN, in general, the predition accuracy should improve when adding more hidden layer and/or hidden nodes. Thus we construct a improved ANN with 2 hidden layers, where there are 19 hidden nodes on the 1st hidden layer and 23 on the 2nd hidden layer. The activation functions are the same.

- Predicting Results Using 12 features
  - Sigmoid function: Prediction Accuracy: 80.7
  - Tanh function: ??
- Here is an Error Rate plot
  Need the 3D plot here
- Here is a ROC plot
- 7 AUC and weighted AUC

§ Feature Selected ANN

Instead of using all 12 features, we decide to use less feature which are relatively independent to each other. We first contructed a feature scatter plot to observe the dependency of features.
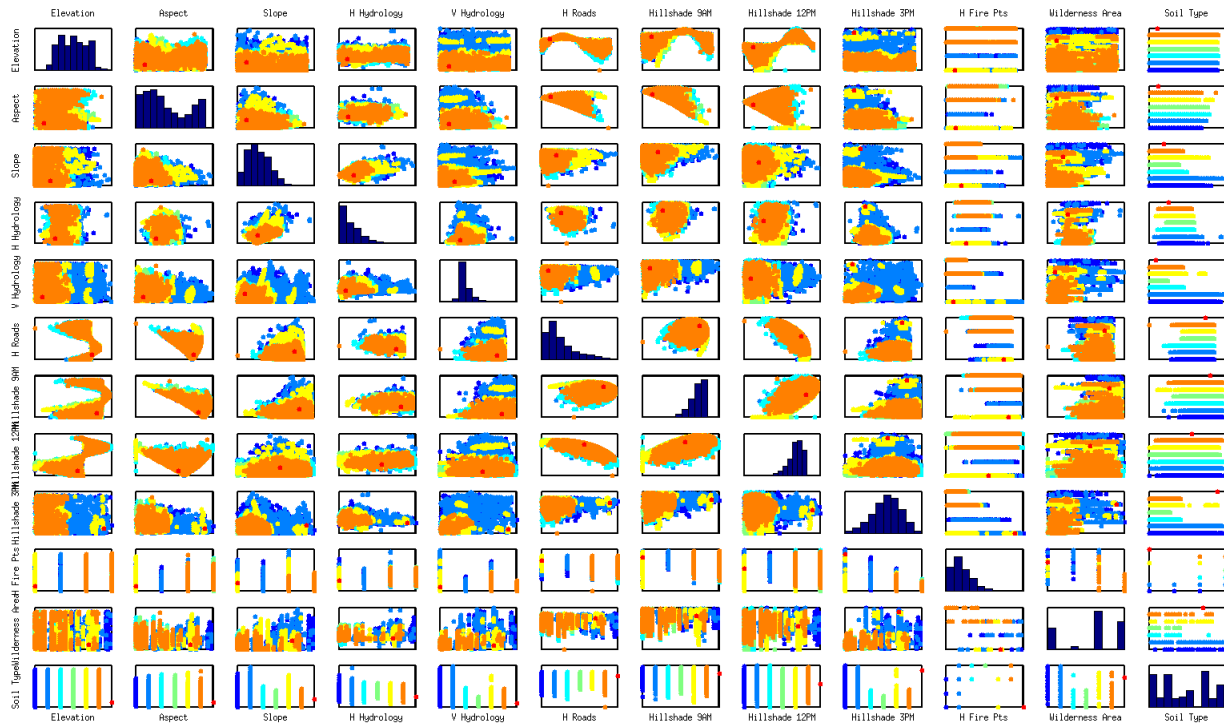
The dependency and correlation of different features are not obvious. So we tested on a few cases, in which the prediction result aren't imporoved.

-
-
-