MACHINE LEARNING WITH .NET

Asmaliza Ahzan @ Emma

asmaliza@iverson.com.my

CLASSIFICATION MODEL

CHAPTER 4

CLASSIFICATION MODELS

- Classification is broken down into two main categories - two-class and multi-class.
- In a two-class classifier, also known as a binary classifier, the prediction simply returns 0 or 1.
- In a multi-class problem, a pre-selected range of return labels, such as virus types or car types, is returned.

BINARY CLASSIFICATION

- A supervised machine learning task that is used to predict which of two classes (categories) an instance of data belongs to.
- The input of a classification algorithm is a set of labeled examples, where each label is an integer of either 0 or 1.
- The output of a binary classification algorithm is a classifier, which you can use to predict the class of new unlabeled instances.

EXAMPLES OF BINARY CLASSIFICATION SCENARIOS

- Understanding sentiment of Twitter comments as either "positive" or "negative".
- Diagnosing whether a patient has a certain disease or not.
- Making a decision to mark an email as "spam" or not.
- Determining if a photo contains a particular item or not, such as a dog or fruit.

BINARY CLASSIFICATION TRAINERS

- AveragedPerceptronTrainer
- SdcaLogisticRegressionBinaryTrainer
- SdcaNonCalibratedBinaryTrainer
- SymbolicSgdLogisticRegressionBinaryTrainer
- LbfgsLogisticRegressionBinaryTrainer
- LightGbmBinaryTrainer
- FastTreeBinaryTrainer
- FastForestBinaryTrainer
- GamBinaryTrainer
- FieldAwareFactorizationMachineTrainer
- PriorTrainer
- LinearSvmTrainer

MULTICLASS CLASSIFICATION

- A supervised machine learning task that is used to predict the class (category) of an instance of data.
- The input of a classification algorithm is a set of labeled examples.
- Each label normally starts as text. It is then run through the TermTransform, which converts it to the Key (numeric) type.
- The output of a classification algorithm is a classifier, which you can use to predict the class of new unlabeled instances.

EXAMPLES OF MULTI-CLASS CLASSIFICATION SCENARIOS

- Categorizing flights as "early", "on time", or "late".
- Understanding movie reviews as "positive", "neutral", or "negative".
- Categorizing hotel reviews as "location", "price", "cleanliness", etc.

MULTICLASS CLASSIFICATION TRAINERS

- LightGbmMulticlassTrainer
- SdcaMaximumEntropyMulticlassTrainer
- SdcaNonCalibratedMulticlassTrainer
- LbfgsMaximumEntropyMulticlassTrainer
- NaiveBayesMulticlassTrainer
- OneVersusAllTrainer
- PairwiseCouplingTrainer

CHOOSING A CLASSIFICATION TRAINER

- Does your problem simply predict a value of true or false, or does it provide a more varied output based on a pre-defined value set?
- If your answer is the former, you need to use a binary classification.
- If the latter, you will need to use a multi-class classification.
- For specific binary classification trainers, SDCA, LightGBM, and FastTree are the most popular options, as well as the most documented.
- For specific multi-class classification trainers, LightGBM and SDCA are the most popular and best-documented options.

CREATING A BINARY CLASSIFICATION APPLICATION

- The application we will be creating is a car-value predictor.
- Given a set of attributes tied to a car, one can predict if the price is a good deal or not.

FASTTREE TRAINER

- FastTree is based on the Multiple Additive Regression Trees (MART) gradient boosting algorithm.
- Gradient boosting is a very popular technique, in which a series of trees are built in a step-wise manner before ultimately selecting the best tree.
- MART takes this approach a step further by learning an ensemble of regression trees that use scalar values in their leaves.
- The FastTree trainer doesn't require normalization but does require all of the feature columns to use a float variable type and the label column to be a bool variable type.

CREATING A MULTI-CLASS CLASSIFICATION APPLICATION

- As mentioned earlier, we will now create a multi-class classification application, categorizing email into one of three categories:
  - Orders
  - Spam
  - Friend
- This multi-class classification application we will be using the SdcaMaximumEntropy trainer.
- The SdcaMaximumEntropy class, as the name implies, is based on the SDCA.

ACTIVITY

Analyze sentiment of website comments with binary classification in ML.Net

THE END