



Kurzdokumentation der Tools

Version 11. Oktober 2013

bug-reports bitte an gunnar.klaemke@phi-tps.de

calldiagfit

Bringt den Netzwerk-Output mit Hilfe eines Fits auf die Diagonale (bezogen auf die Korrelation zum Target), so dass der transformierte Netzwerk-Output direkt als Wahrscheinlichkeit interpretiert werden kann.

Aufrufkonvention

Das Programm wird zunächst als Teacher,

```
calldiagfit <csvInFile> <expertiseFile> <sample> teacher <est_pred_name>  
            <truth_name> [< weight_name>]
```

und dann als Expert aufgerufen.

```
calldiagfit <csvInFile> <expertiseFile> <sample> expert <est_pred_name>  
            <transformed_pred_name>
```

- **csvInFile**: Eingabe-Datei im .csv Format
- **expertiseFile**: Name der Datei, in der die Expertise abgelegt wird (Teacher-Modus), bzw. von der die Expertise gelesen wird (Expert-Modus).
- **sample**: Der Index des Samples bei Trainings mit cross-referencing. Kann ansonsten auf einen beliebigen Wert ≤ 0 gesetzt werden.
- **est_pred_name**: Name der Variablen, die den Netzwerk-Output beinhaltet.
- **truth_name**: Name der Variablen, die die truth bzw. das target beinhaltet (nur im Teacher-Modus).
- **transformed_pred_name**: Name der Variablen, in die der transformierte NB-Output geschrieben werden soll (nur im Expert-Modus)
- **weight_name** (optional): Name der Variablen, die das Ereignisgewicht angibt (nur im Teacher-Modus).

Der Expert legt die transformierte Größe zusammen mit den Input-Daten in einer Datei mit Namen 'diagfit_<csvInFile>' im Verzeichnis der Input-Datei ab.

call2dboost

Wendet einen 2d-boost auf die Daten des input-files an. Beim 2d-boost wird die Korrelation der 2-dimensionalen Verteilungen von jeweils einem Paar von Input-Variablen zum Target ausgenutzt.

Aufrufkonvention

Das Programm wird zunächst als Teacher,

```
call2dboost <csvInFile> <expertiseFile> <sample> <needed_varnames>  
           teacher <estimated_prediction_name> <target_name> [<weight_name>]
```

und dann als Expert aufgerufen.

```
call2dboost <csvInFile> <expertiseFile> <sample> <needed_varnames>  
           expert <estimated_prediction_name> <pred_2dboost_name>
```

- **csvInFile**: Eingabe-Datei im .csv Format
- **expertiseFile**: Name der Datei, in der die Expertise abgelegt wird (Teacher-Modus), bzw. von der die Expertise gelesen wird (Expert-Modus).
- **sample**: Der Index des Samples bei Trainings mit cross-validation. Kann ansonsten auf einen beliebigen Wert ≤ 0 gesetzt werden.
- **needed_varnames**: Komma-separierte Liste der Input-Variablen, welche zum 2dboost verwendet werden sollen. Es sollten wenigstens zwei Variablenamen angegeben werden.
- **est_pred_name**: Name der Variablen, die den Netzwerk-Output (vor dem 2dboost) beinhaltet.
- **target_name**: Name der Variablen, die die truth bzw. das target beinhaltet (nur im Teacher-Modus).
- **pred_2dboost_name**: Name der Variablen, in die der geboostete NB-Output geschrieben werden soll (nur im Expert-Modus)
- **weight_name** (optional): Name der Variablen, die das Ereignisgewicht angibt (nur im Teacher-Modus).

Der Expert legt die transformierte Größe zusammen mit den Input-Daten in einer Datei mit Namen '2dboost_<csvInFile>' im Verzeichnis der Input-Datei ab.

plotgini

Generiert einige Standard-Plots im .ps-Format (Target-Verteilung, Diagonal-Plot, Gini-Index, u.a.). Weitergehende Erläuterungen zu den Plots sind in “docginiplot.ps” zu finden.

Aufrufkonvention

Es gibt zwei Aufruf-Modi: einen für Klassifizierungs-Trainings und einen für Density-Trainings.

für Klassifizierungen

```
plotgini [options] <rootInFile> <outputfile> <truth_name>  
        <est_pred_names, separated by commas> [<weight_name>]
```

- **options:**

- -C/-c muss angegeben werden
- -l/-L für die lange Fassung inkl. *S/B*-Scans.
- -s/-S für die Kurzform (default).
- -F/-f falls die Variablen im Root-Tree als 'float' gespeichert sind (die Default-Annahme ist 'double')

- **rootInFile:** Eingabe-Datei im .root Format
- **outputFile:** Name der Ausgabe-Datei. Es wird outputFile.ps erzeugt.
- **truth_name:** Name der Variable mit der truth- bzw. target-Information.
- **est_pred_names:** eine oder mehrere Variablennamen (als Komma-separierte Liste), die jeweils einen Network-Output bezeichnen.
- **weight_name** (optional): Name der Variable, die das Ereignis-Gewicht beschreibt.

für Densities

```
plotgini [options] <rootInFile> <outputfile> <truth_names, separated by commas>  
        <mean_names, separated by commas>  
        <median_names, separated by commas> [<weight_name>]
```

- **options:**

- -D/-d muss angegeben werden
- -F/-f falls die Variablen im Root-Tree als 'float' gespeichert sind (die Default-Annahme ist 'double')

- **rootInFile:** Eingabe-Datei im .root Format
- **outputFile:** Name der Ausgabe-Datei. Es wird outputFile.ps erzeugt.

- **truth_names**: eine oder mehrere Variablennamen (als Komma-separierte Liste), die die Truth- bzw. das Target beschreiben.
- **mean_names**: Liste von Variablennamen, die den Mean der Netzwerk-Vorhersage angeben. Die Reihenfolge muss der Liste der angegebenen truth-Variablen entsprechen.
- **median_names**: Liste von Variablennamen, die den Median der Netzwerk-Vorhersage beschreiben. Die Reihenfolge muss ebenfalls der Liste der angegebenen truth-Variablen entsprechen.
- **weight_name** (optional): Name der Variable, die das Ereignis-Gewicht beschreibt.