A Matlab Toolbox for Comparing Supervised Classification Algorithms

Olcay Taner Yıldız

Technical Report, FBE/COE-04/2004-20

Institute of Graduate Studies in Science and Engineering

Department of Computer Engineering

Boğaziçi University

yildizol@cmpe.boun.edu.tr

August 31, 2004

TABLE OF CONTENTS

1. Distribution and Inverse Distribution Functions

1.1. Chi-Square Distribution

Usage: probability value = chi(x, freedom)

Parameters:

- **x**: The critical value of the chi-square distribution.
- **freedom**: The degree of freedom of the chi-square distribution.

Return: The probability of the chi-square distribution with critical value of x and with degree of freedom freedom.

Range: Returns 1 if x $<$ 0 or freedom $<$ 1

Example: a = chi(2.56, 3)

1.2. Inverse Chi-Square Distribution

Usage: critical value = chiinv(p, freedom)

Parameters:

- **p**: The probability value of the chi-square distribution.
- **freedom**: The degree of freedom of the chi-square distribution.

Return: The critical value of the chi-square distribution with probability value of p and with degree of freedom freedom.

Range: Returns 0 if p $\leq$ 0, 99999 if p $\geq$ 1

Example: a = chiinv(0.99, 5)

1.3. F Distribution

Usage: probability value = fdist(F, freedom1, freedom2)

Parameters:

- **F**: The critical value of the F distribution.
- **freedom1**: The first degree of freedom of the F distribution.
- **freedom2**: The second degree of freedom of the F distribution.

Return: The probability of the F distribution with critical value of F and with degree of freedoms freedom1 and freedom2.

Range: Returns 1 if F $<$ 0.000001 or when either of the freedoms $<$ 1

Example: a = fdist(2.56, 5, 7)

1.4. Inverse F Distribution

Usage: critical value = fdistinv(p, freedom1, freedom2)

Parameters:

- **p**: The probability value of the F distribution.
- **freedom1**: The first degree of freedom of the F distribution.
- **freedom2**: The second degree of freedom of the F distribution.

Return: The critical value of the F distribution with probability value of p and with degree of freedoms freedom1 and freedom2.

Range: Returns 0 if p $\leq$ 0 or $\geq$ 1

Example: a = fdistinv(0.95, 10, 5)

1.5. Studentized Range Distribution

Usage: probability value = qdist(q, v, r)

Parameters:

- **q**: The critical value of the studentized range distribution.
- **v**: The first degree of freedom of the studentized range distribution.
- **r**: The second degree of freedom of the studentized range distribution.

Return: The probability of the studentized range distribution with critical value of q and with degree of freedoms v and r.

Example: a = qdist(4.05, 3, 6)

1.6. Inverse Studentized Range Distribution

Usage: critical value = qdistinv(p, v, r)

Parameters:

- **p**: The probability value of the studentized range distribution.
- **v**: The first degree of freedom of the studentized range distribution.
- **r**: The second degree of freedom of the studentized range distribution.

Return: The critical value of the studentized range distribution with probability value of p and with degree of freedoms v and r.

Example: a = qdistinv(0.999, 6, 4)

1.7. t Distribution

Usage: probability value = chi(T, freedom)

Parameters:

- **T**: The critical value of the t distribution.
- **freedom**: The degree of freedom of the t distribution.

Return: The probability of the t distribution with critical value of T and with degree of freedom freedom.

Range: Returns 1 if degree of freedom < 1

Example: a = tdist(-1.23, 5)

1.8. Inverse t Distribution

Usage: critical value = tdistinv(p, freedom)

Parameters:

- **p**: The probability value of the t distribution.
- **freedom**: The degree of freedom of the t distribution.

Return: The critical value of the t distribution with probability value of p and with degree of freedom freedom.

Range: Returns 0 if $p \leq 0$ or $p \geq 1$

Example: a = tdistinv(0.001, 5)

1.9. Standardized Normal Distribution

Usage: probability value = znormal(z)

Parameters:

- **z**: The critical value of the standardized normal distribution.

Return: The probability of the standardized normal distribution with critical value of z.

Range: Returns 0.5 if z = 0.

Example: a = znormal(-5.367)

1.10. Inverse Standardized Normal Distribution

Usage: critical value = zinverse(p)

Parameters:

- **p**: The probability value of the standardized normal distribution.

Return: The critical value of the standardized normal distribution with probability value of p.

Range: Returns 0 if $p \leq 0$ or ≥ 1

Example: a = zinverse(0.05)

2. Classifier Comparison Functions

2.1. Multiple Tests

For all multiple tests given below, the examples are given on the data matrix

```
A4 = [9 15 18 21
    11 13 16 19
    10 17 15 17
    13 14 17 23
     8 15 13 18
    12 16 15 20
    10 15 14 16
    11 17 17 21
     7 14 15 21
    10 13 16 19]
```

2.1.1. Anova Test

Usage: accepted = anova(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where i'th column represents the error rates of the i'th classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2 = \ldots = m_k$.

Return: 1 if the null hypothesis is accepted (all classifiers have the same mean), 0 if

the null hypothesis is rejected (at least one of the classifiers have different error rate from at least one of the classifiers).

Example: anova(A4, 0.95) returns 0.

2.1.2. Kruskal–Wallis Test

Usage: accepted = kruskal(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where i'th column represents the error rates of the i'th classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2 = \ldots = m_k$.

Return: 1 if the null hypothesis is accepted (all classifiers have the same mean), 0 if the null hypothesis is rejected (at least one of the classifiers have different error rate from at least one of the classifiers).

Example: kruskal(A4, 0.99) returns 0.

2.1.3. Van der Waerden Test

Usage: accepted = waerdan(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where i'th column represents the error rates of the i'th classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2 = \ldots = m_k$.

Return: 1 if the null hypothesis is accepted (all classifiers have the same mean), 0 if the null hypothesis is rejected (at least one of the classifiers have different error rate from at least one of the classifiers).

Example: waerdan(A4, 0.90) returns 0.

2.1.4. Cochran's Test

Usage: accepted = cochran(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where the element on the i'th column and j'th row represents the performance of the i'th classifier on the j'th instance on the test set. It is 1, if the instance is correctly classified, 0 if the instance is incorrectly classified.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2 = \ldots = m_k$.

Return: 1 if the null hypothesis is accepted (all classifiers have the same mean), 0 if the null hypothesis is rejected (at least one of the classifiers have different error rate from at least one of the classifiers).

Example: c = cochran(B4, 0.95).

2.2. Pairwise Tests

For all pairwise tests given below, the examples are given on the data matrix

```
A12 = [9 15
    11 13
    10 17
    13 14
     8 15
    12 16
    10 15
    11 17
     7 14
    10 13 ]
```

2.2.1. Bonferroni Test

Usage: accepted = bonferroni(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where the first column contains the error rates of the first classifier and second column contains the error rates of the second classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2$.

Return: 1 if the null hypothesis is accepted (both classifiers have the same mean), 0 if the null hypothesis is rejected (two classifiers have different error rates).

Example: bonferroni(A12, 0.95) returns 0.

2.2.2. 5× 2 cv F Test

Usage: accepted = ftest5x2(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where the first column contains the error rates of the first classifier and second column contains the error rates of the second classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2$.

Return: 1 if the null hypothesis is accepted (both classifiers have the same mean), 0 if the null hypothesis is rejected (two classifiers have different error rates).

Example: ftest5x2(A12, 0.95) returns 1.

2.2.3. Mc Nemar's Test

Usage: accepted = mcnemar(results, confidencelevel)

Parameters:

- **results**: Two by two matrix.

  ```
  [a b
   c d]
  ```

  a is the number of instances, which both classifiers have correctly classified. d is the number of instances, which both classifiers have incorrectly classified. b is the number of instances, which first classifier has correctly but second classifier has incorrectly classified. c is the number of instances, which first classifier has incorrectly but second classifier has correctly classified.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2$.

Return: 1 if the null hypothesis is accepted (both classifiers have the same mean), 0 if the null hypothesis is rejected (two classifiers have different error rates).

Example: mcnemar([63, 21; 4, 12], 0.95) returns 0.

2.2.4. Scheffe Test

Usage: accepted = scheffe(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where the first column contains the error rates of the first classifier and second column contains the error rates of the second classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2$.

Return: 1 if the null hypothesis is accepted (both classifiers have the same mean), 0 if the null hypothesis is rejected (two classifiers have different error rates).

Example: scheffe(A12, 0.95) returns 0.

2.2.5. Sign Test

Usage: accepted = signtest(plus, minus, confidencelevel)

Parameters:

- **plus**: Total number of positive examples.
- **minus**: Total number of negative examples.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : O(+) = O(-)$.

Return: 1 if the null hypothesis is accepted (there is not significant difference between positive and negative example counts), 0 if the null hypothesis is rejected (there is significant difference between positive and negative example counts).

Example: signtest(13, 17, 0.95) returns 1.

2.2.6. 5× 2 cv t Test

Usage: accepted = ttest5x2(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where the first column contains the error rates of the first classifier and second column contains the error rates of the second classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2$.

Return: 1 if the null hypothesis is accepted (both classifiers have the same mean), 0 if the null hypothesis is rejected (two classifiers have different error rates).

Example: ttest5x2(A12, 0.95) returns 1.

2.2.7. K Fold Crossvalidated Paired t Test

Usage: accepted = ttestcv(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where the first column contains the error rates of the first classifier and second column contains the error rates of the second classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2$.

Return: 1 if the null hypothesis is accepted (both classifiers have the same mean), 0 if the null hypothesis is rejected (two classifiers have different error rates).

Example: ttestcv(A12, 0.95) returns 0.

2.2.8. Tukey Test

Usage: accepted = tukey(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where the first column contains the error rates of the first classifier and second column contains the error rates of the second classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2$.

Return: 1 if the null hypothesis is accepted (both classifiers have the same mean), 0 if the null hypothesis is rejected (two classifiers have different error rates).

Example: tukey(A12, 0.95) returns 0.

2.2.9. Two sample t Test

Usage: accepted = twosamplet(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where the first column contains the error rates of the first classifier and second column contains the error rates of the second classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2$.

Return: 1 if the null hypothesis is accepted (both classifiers have the same mean), 0 if the null hypothesis is rejected (two classifiers have different error rates).

Example: twosamplet(A12, 0.95) returns 0.

2.2.10. Wilcoxon Test

Usage: accepted = wilcoxon(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where the first column contains the error rates of the first classifier and second column contains the error rates of the second classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2$.

Return: 1 if the null hypothesis is accepted (both classifiers have the same mean), 0

if the null hypothesis is rejected (two classifiers have different error rates).

Example: wilcoxon(A12, 0.95) returns 0.

2.3. Range Tests

In each of the range tests below, we assume that the error rates of the classifiers are in increased order.

2.3.1. Newman–Keuls Test

Usage: equalities = newmankeuls(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where i'th column represents the error rates of the i'th classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2 = \ldots = m_k$.

Return: an array of equal classifier indexes. [1 3] means that classifiers 1, 2, 3 are equal to each other. [1, 2; 2, 5] means that classifiers 1, 2 are equal to each other. Also classifiers 2, 3, 4, 5 are equal to each other.

Example: newmankeuls(A4, 0.95) returns [2 3].

2.3.2. Duncan Test

Usage: equalities = duncan(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where i'th column represents the error rates of the i'th classifier.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: $H_0 : m_1 = m_2 = \ldots = m_k$.

Return: an array of equal classifier indexes. [1 3] means that classifiers 1, 2, 3 are equal to each other. [1, 2; 2, 5] means that classifiers 1, 2 are equal to each other. Also classifiers 2, 3, 4, 5 are equal to each other.

Example: duncan(A4, 0.95) returns [2 3].

2.4. Goodness of Fit Test

Usage: accepted = goodnessoffit(treatments, confidencelevel)

Parameters:

- **treatments**: Two dimensional matrix, where observed examples are stored in the first column, expected example counts are stored in the second column.
- **confidencelevel**: acceptance level of the null hypothesis.

Null Hypothesis: H_0: P(x is in group j) = p_j for all $j = 1, \ldots, c$.

Return: 1 if the null hypothesis is accepted, 0 if the null hypothesis is rejected.

Example: goodnessoffit(C2, 0.95).