EFC2 - Exercise 2

Rafael Claro Ito

September 2019

1 Source files

All code presented and all figures showed here can be found at the following GitHub repository: `https://github.com/ito-rafael/IA006C-MachineLearning/tree/master/efc2`
In this repository, one can found the following files:

- Jupyter Notebook
  - efc2_pre-ex1.ipynb
  - efc2_ex1_binary_classification.ipynb
  - efc2_ex2_multiclass_classification.ipynb
  - efc2_ex2_knn.ipynb
- LaTeX
  - efc2.tex

The notebook "efc2_pre-ex1" plots the histograms for the exercise 1 and it is mainly used for data visualization. It shows the input features histograms for the raw data and also after a data stardardization.

The notebook "efc2_ex1_binary_classification" effectively implements the logistic regression used to perform a binary classification.

The notebooks "efc2_ex2_multiclass_classification" and "efc2_ex2_knn" implements the algorithms to perform a multiclass classification proposed in exercise 2. The former one uses the softmax approach while latter one implements the K-Nearest Neighbors (KNN) algorithm.

2 Part 1 - Binary Classification

2.1 a) Input features characteristics analysis considering the histograms and correlation measures between them.

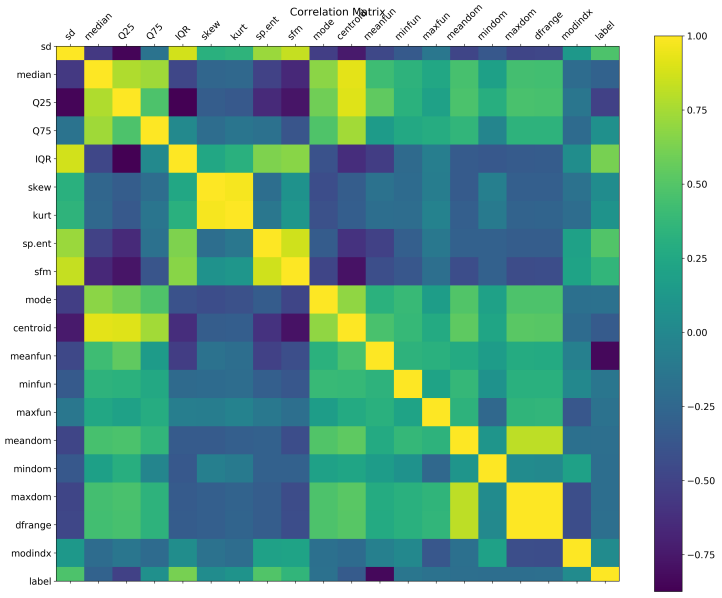
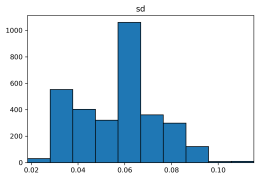
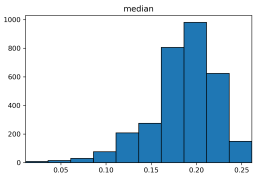


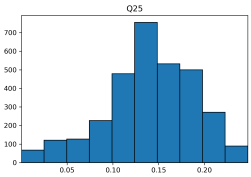
Figure 4: Correlation between input features



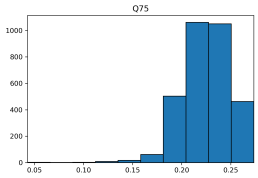
(a) sd



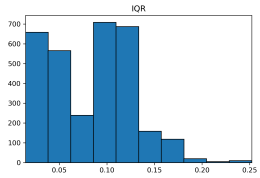
(b) median



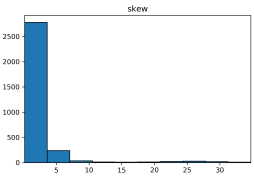
(c) Q25



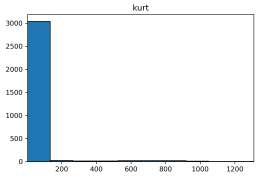
(d) Q75



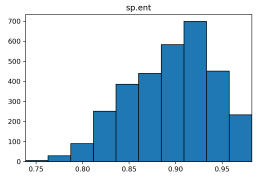
(e) IQR



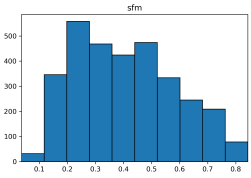
(f) skew



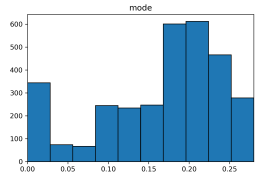
(g) kurt



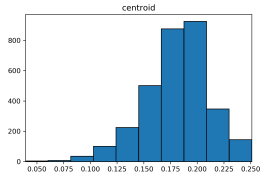
(h) sp ent



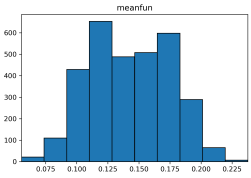
(i) smf



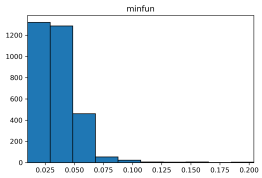
(j) mode



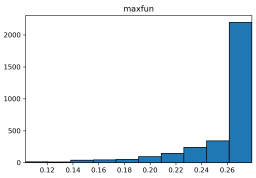
(k) centroid



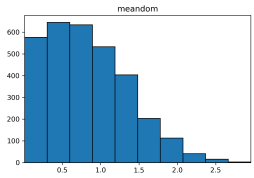
(l) meanfun



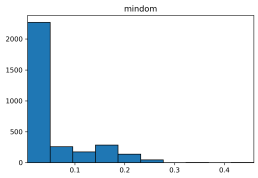
(m) minfun



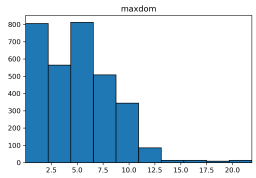
(n) maxfun



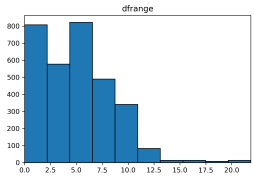
(o) meandom



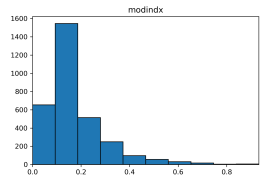
(p) mindom



(q) maxdom

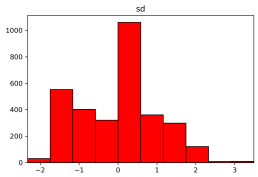


(r) dfrange

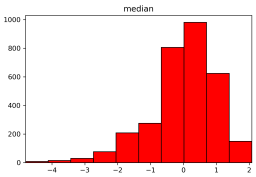


(s) modindx

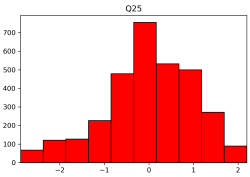
Figure 2: Histograms of raw input features



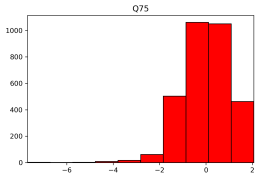
(a) sd



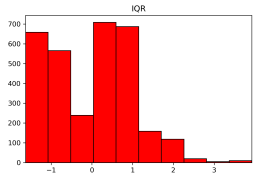
(b) median



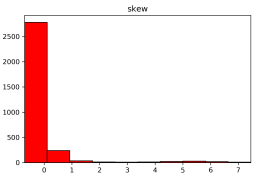
(c) Q25



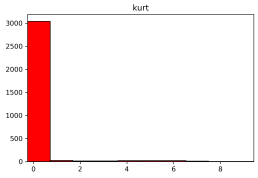
(d) Q75



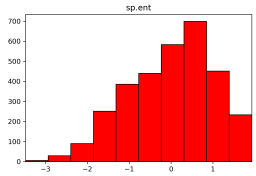
(e) IQR



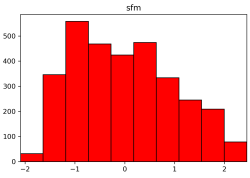
(f) skew



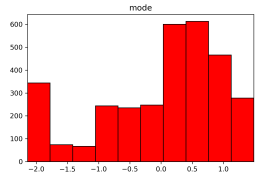
(g) kurt



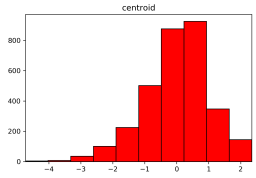
(h) sp ent



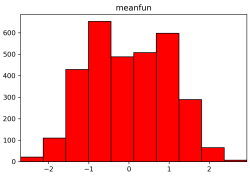
(i) smf



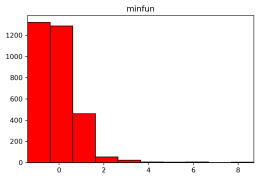
(j) mode



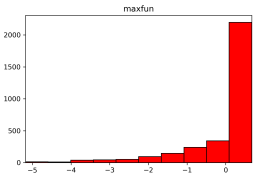
(k) centroid



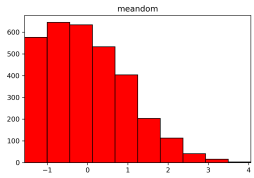
(l) meanfun



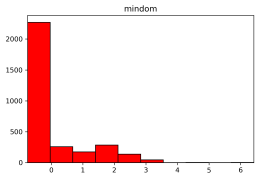
(m) minfun



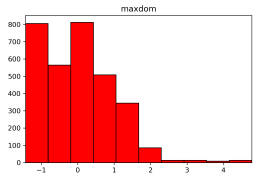
(n) maxfun



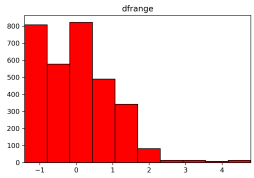
(o) meandom



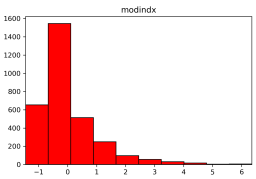
(p) mindom



(q) maxdom



(r) dfrange



(s) modindx

Figure 3: Histograms of input features after data standardization

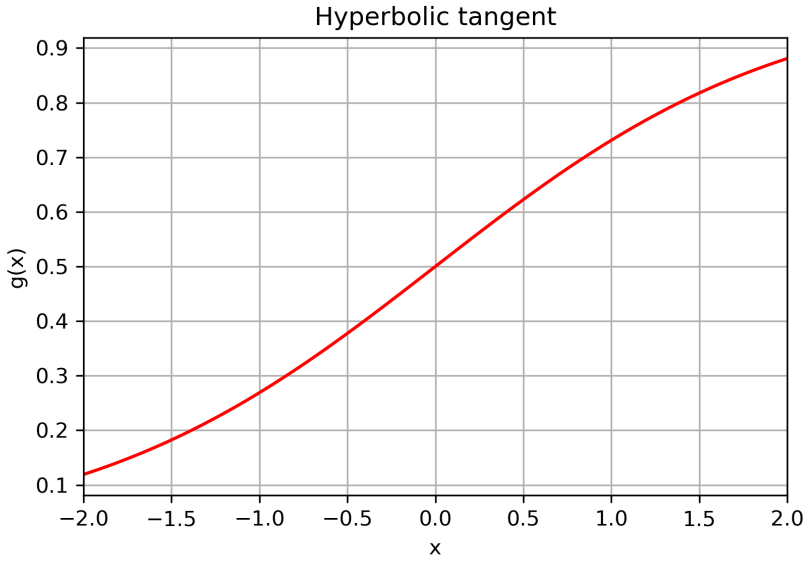


Figure 5: Logistic function

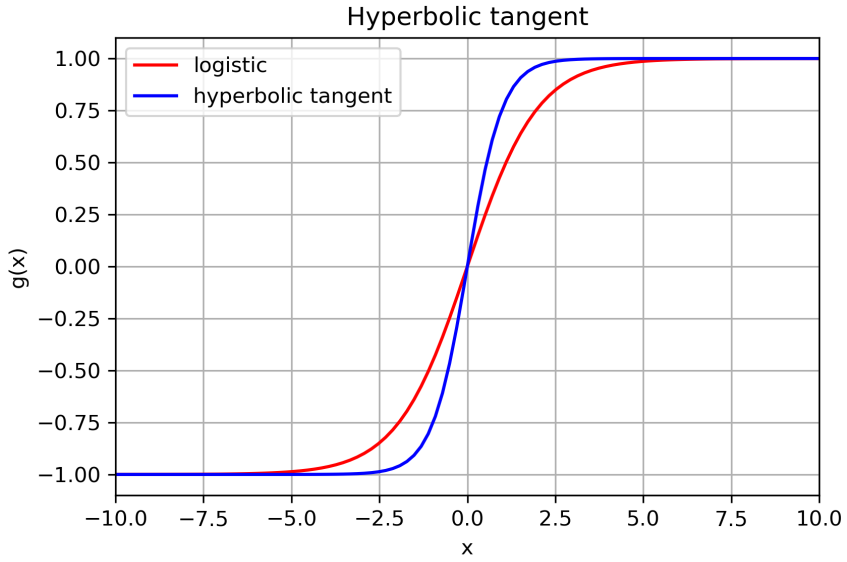


Figure 6: Comparasion between hyperbolic tangent and logistic function(scaled and with offset)

The first thing we did before starting this exercise was to plot all the data, as can be seen in figure 5.

2.2 b)

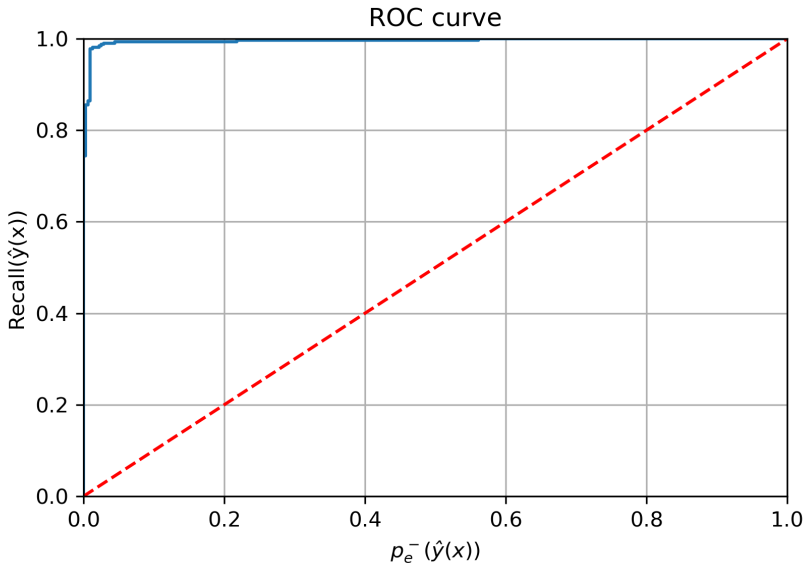


Figure 7: ROC curve

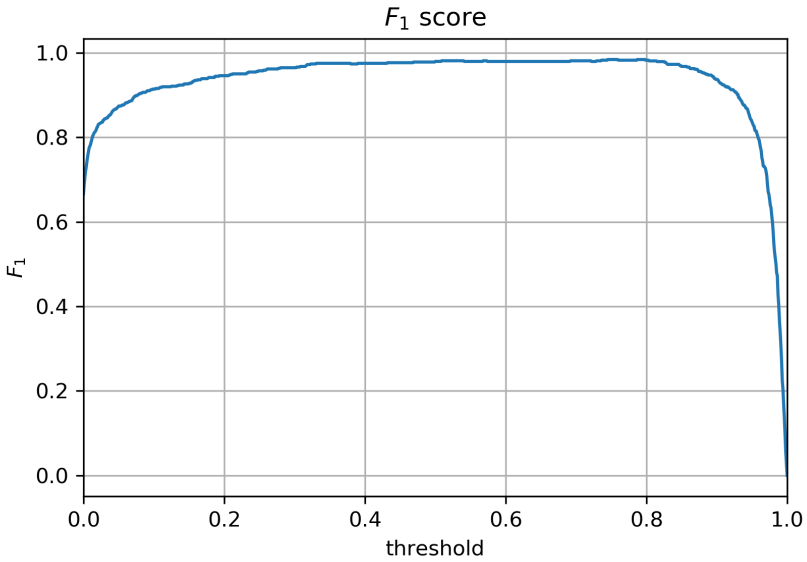


Figure 8: F1-score curve

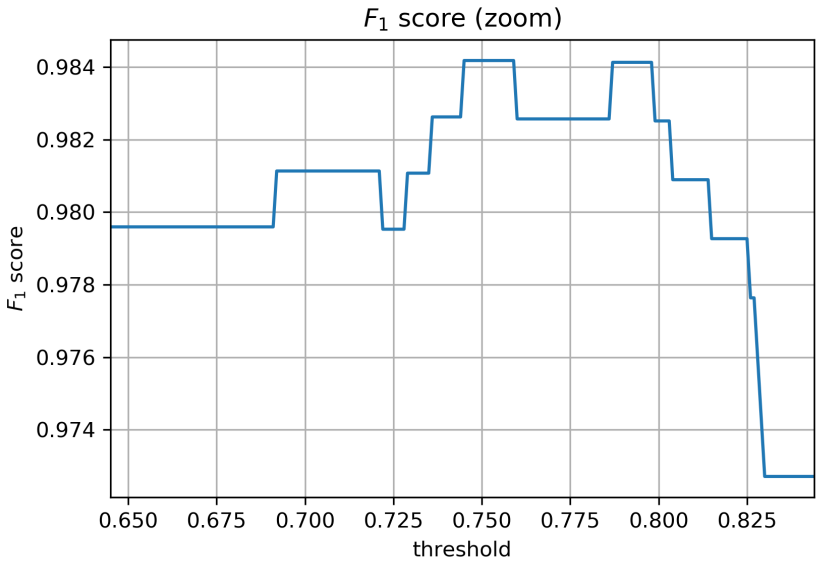


Figure 9: Zoom in F1-score curve

2.3 c)

3 Part 2 - Multiclass Classification

3.1 a)

output ¡br¿ 1 – caminhada ¡br¿ 2 – subindo escadas ¡br¿ 3 – descendo escadas ¡br¿ 4 – sentado ¡br¿ 5 – em pé ¡br¿ 6 – deitado ¡br¿ ¡br¿ **one-hot enconding** ¡br¿ $[1\ 0\ 0\ 0\ 0\ 0]^T$: walking ¡br¿ $[0\ 1\ 0\ 0\ 0\ 0]^T$: climbing stairs ¡br¿ $[0\ 0\ 1\ 0\ 0\ 0]^T$: going down stairs ¡br¿ $[0\ 0\ 0\ 1\ 0\ 0]^T$: seated ¡br¿ $[0\ 0\ 0\ 0\ 1\ 0]^T$: standing ¡br¿ $[0\ 0\ 0\ 0\ 0\ 1]^T$: lying ¡br¿

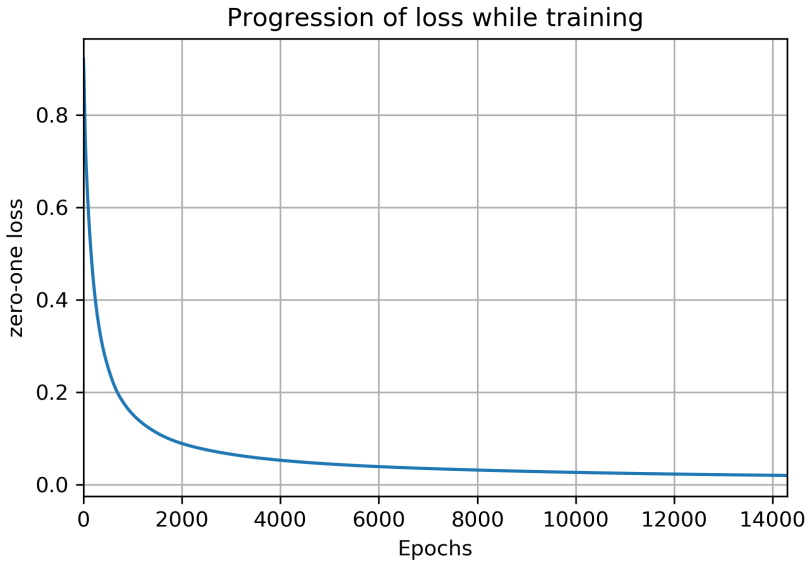


Figure 10: Zero-one loss curve

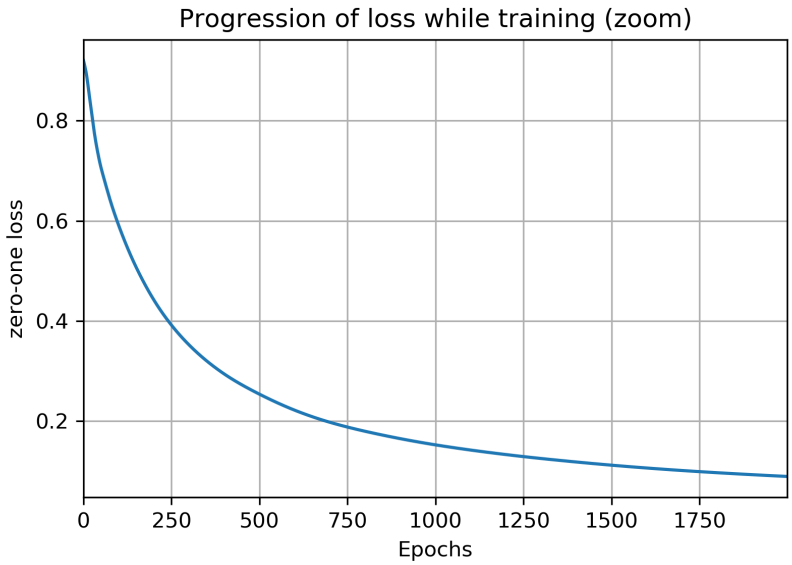


Figure 11: Zoom in zero-one loss curve

3.2 b)

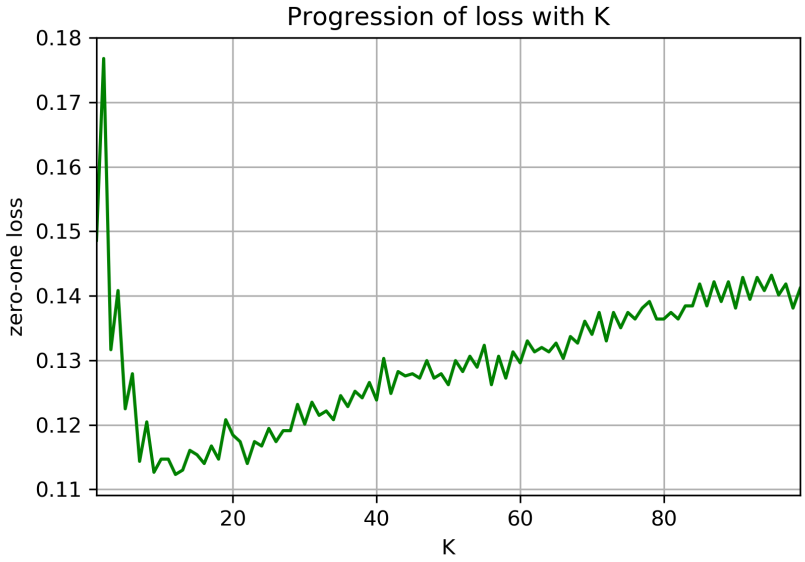


Figure 12: Zero-one loss curve (uniform weights

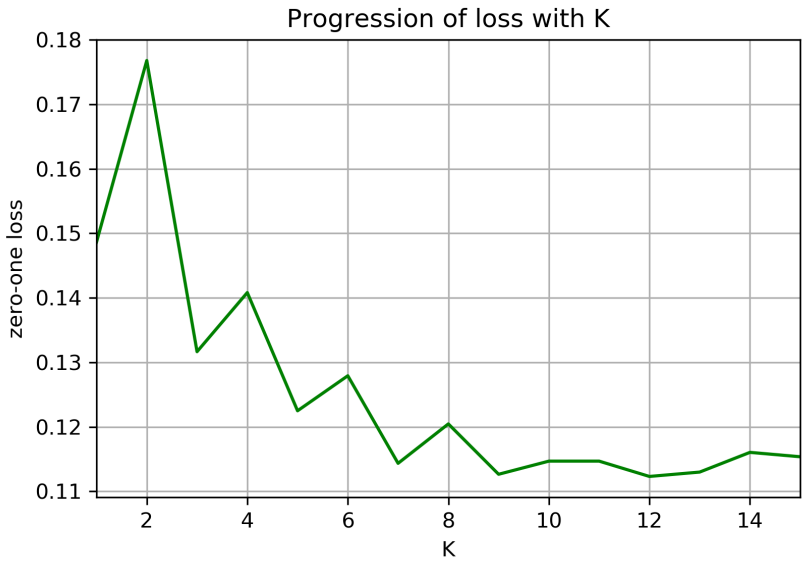


Figure 13: Zero-one loss curve (uniform weights

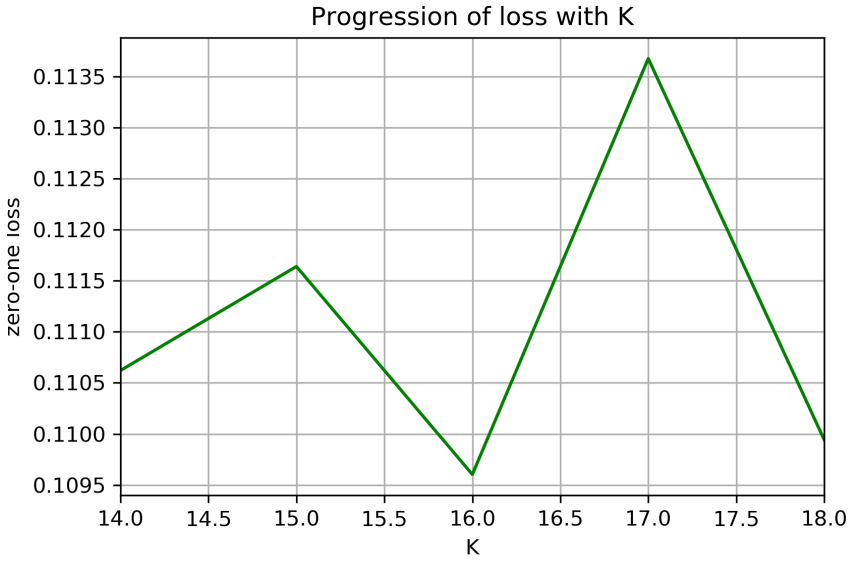


Figure 14: Zero-one loss curve (uniform weights

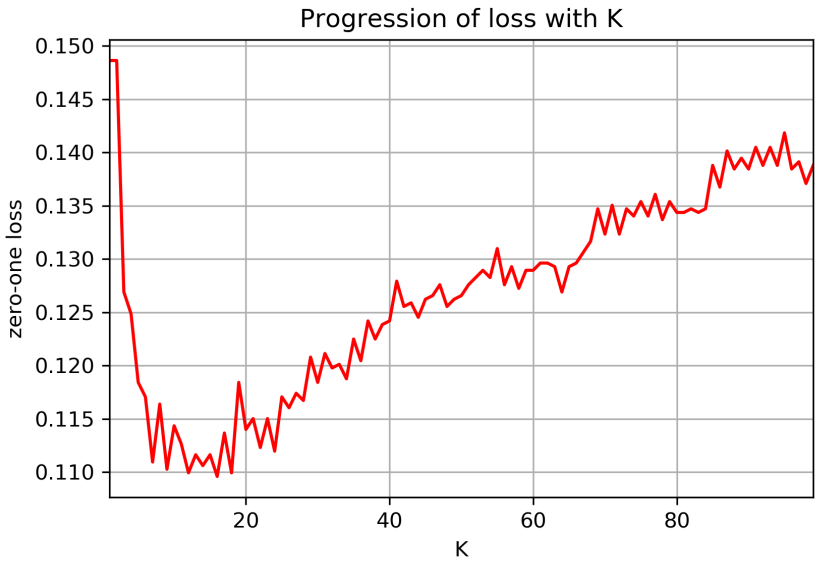


Figure 15: Zero-one loss curve (inversely proportional to distance weights

Figure 16: Zero-one loss curve (inversely proportional to distance weights

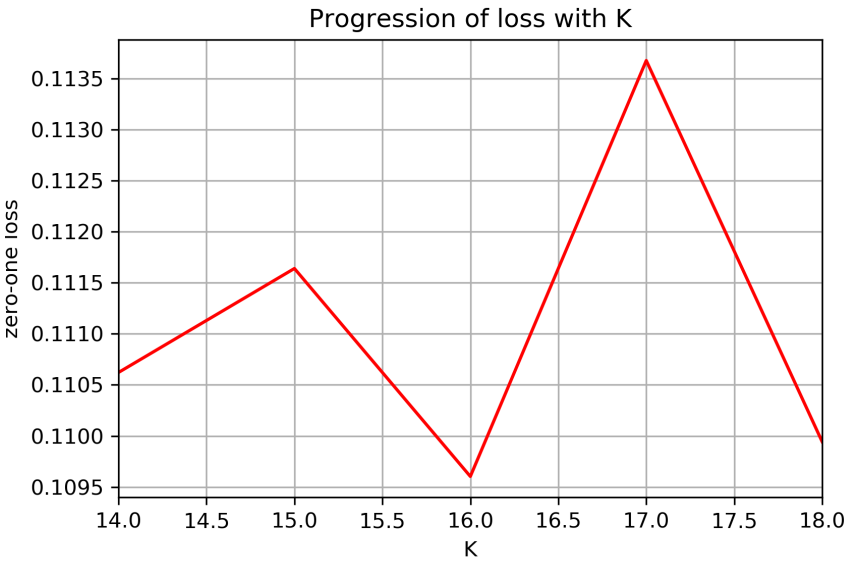


Figure 17: Zero-one loss curve (inversely proportional to distance weights