University of Stuttgart | ANALYTIC COMPUTING

# Machine Learning (SS 24)

## Assignment 2: Classification

Akram Sadat Hosseini
Akram.Hosseini@ki.uni-stuttgart.de

Nadeen Fathallah
Nadeen.Fathallah@ki.uni-stuttgart.de

Tim Schneider
tim.schneider@ki.uni-stuttgart.de

Osama Mohammed
osama.mohammed@ki.uni-stuttgart.de

Daniel Frank
daniel.frank@ki.uni-stuttgart.de

Submit your solution in ILIAS as a single PDF file.[1] Make sure to list the full names of all participants, matriculation number, study program, and B.Sc. or M.Sc on the first page. Optionally, you can *additionally* upload source files (e.g. PPTX files). If you have any questions, feel free to ask them in the exercise forum in ILIAS.

**Submission is open until Monday, 29.04.2024, 12:00 noon.**

[1] Your drawing software probably allows exporting as PDF. An alternative option is to use a PDF printer. If you create multiple PDF files, use a merging tool (like pdfarranger) to combine the PDFs into a single file.

# Task 1: Data Labels (31 Points)

Consider the fruit classification example from the lecture with the categories $\{$*banana*, *grapes*, *apple*, *orange*$\}$.

1. **Task (8 Points)** Formally define a dataset $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$ that consists of
   - a banana, that is *yellow*, *curved*, *smooth*,
   - an apple that is *red*, *round*, *smooth*,
   - an orange that is *orange*, *round*, *rough*,
   - a banana, that is *brown*, *curved*, *smooth*,
   - grapes, that are *green*, *oval*, *smooth*,
   - an apple that is *green*, *round*, *smooth*.

   Explicitly provide $\mathcal{D}$, $\mathcal{X}$ and $\mathcal{Y}$.
2. **Task (6 Points)** Associated with each class is a label $Y = \{l_0, l_1, l_2, l_3\}$ that needs to be defined. Assume you assigned the labels $l_0 = 0 \in \mathbb{R}$, $l_1 = 1 \in \mathbb{R}$, $l_2 = 2 \in \mathbb{R}$, and $l_3 = 3 \in \mathbb{R}$ for the fruit categories in their given order. To judge the quality of a Machine Learning algorithm, we need a performance measure. As in the lecture we use the Euclidean distance metric $\|\cdot\|_2$. Compute it between all pairs of predicted labels $\hat{y} \in \mathcal{Y}$ and true labels $y \in \mathcal{Y}$. Why might this labeling be problematic?
3. **Task (3 Points)** Compute the loss of the entire dataset for the classifier

$$\hat{f}(x_i) = \begin{cases} 0 & \text{, if } x_{i,0} = \text{yellow} \\ 1 & \text{, if } x_{i,0} = \text{green} \\ 2 & \text{, if } x_{i,0} = \text{red} \\ 3 & \text{, else} \end{cases} \tag{1}$$

4. **Task (6 Points)** Come up with a different labeling $y \in \mathbb{R}^m$ such that the euclidean distance $\|\cdot\|_2$ between all pairs of classes is uniform.
5. **Task (3 Points)** Again, compute the loss for the entire dataset.
6. **Task (5 Points)** Many algorithms you will learn about in the lecture will also require a numerical representation of the input $\mathcal{X}$ as well. Provide a single vector $x_i \in \mathbb{R}^m$ for each data point with dimension $m \in \mathbb{N}$ of your choice.

# Task 2: Classification Metrics (38 Points)

## Task 2.1: $F_1$-Score (18 Points)

1. **Task (2 Points)** What's the *minimum* and *maximum* possible $F_1$-score? Shortly describe a classifier that achieves that score in a binary classification task.
2. **Task (8 Points)** In the lecture we defined $F_1 = \frac{2 \cdot \text{Prec} \cdot \text{Rec}}{\text{Prec} + \text{Rec}}$. What's the $F_1$ in case of $\text{Prec} = \text{Rec} = 0$? *(Hint: $\frac{0}{0}$ is not defined.)*
3. **Task (8 Points)** Let $\mathcal{D}$ be a dataset with $|\mathcal{D}| = n$, two different classes ($|\mathcal{Y}| = 2$) and $|\{(x_i, y_i) \in \mathcal{D} | y_i = \textit{positive}\}| = m$.
   - What's the expected $F_1$-score ($\mathbb{E}[F_1]$) of a classifier that uniformly picks a class at random?
   - What's the expected $F_1$-score of the former with $n - m = m$?

## Task 2.2: Confusion Matrices (20 Points)

Based on the following confusion matrices, which classifier would you choose? Justify your answer since there might not be a single valid decision in some scenarios. Would you choose the classifier with confusion matrix $C_1$ or the classifier with $C_2$ ...

1. **Task (4 Points)** ... for an arbitrary classification task?

$$C_1 = \begin{pmatrix} 250 & 0 \\ 0 & 250 \end{pmatrix} \qquad C_2 = \begin{pmatrix} 125 & 125 \\ 125 & 125 \end{pmatrix}$$

2. **Task (4 Points)** ... for an arbitrary classification task?

$$C_1 = \begin{pmatrix} 0 & 250 \\ 250 & 0 \end{pmatrix} \qquad C_2 = \begin{pmatrix} 125 & 125 \\ 125 & 125 \end{pmatrix}$$

3. **Task (6 Points)** ... for fire detection task? (positive class means fire present)

$$C_1 = \begin{pmatrix} 200 & 25 \\ 75 & 200 \end{pmatrix} \qquad C_2 = \begin{pmatrix} 200 & 75 \\ 25 & 200 \end{pmatrix}$$

4. **Task (6 Points)** ... for deciding which mushrooms in the forest to pick for your soup? (positive class means edible)

$$C_1 = \begin{pmatrix} 200 & 25 \\ 75 & 200 \end{pmatrix} \qquad C_2 = \begin{pmatrix} 200 & 75 \\ 25 & 200 \end{pmatrix}$$

# Task 3: Extended Confusion Martices (31 Points)

Given a classifier predicts the following in a multi-class classification problem.

| $\hat{y}$ (Predicted) | $y$ (Actual Label) |
|---|---|
| Apple | Apple |
| Apple | Apple |
| Orange | Grapes |
| Grapes | Grapes |
| Orange | Orange |
| Apple | Grapes |
| Orange | Orange |
| Grapes | Orange |
| Grapes | Grapes |
| Grapes | Grapes |
| Orange | Apple |
| Grapes | Grapes |
| Apple | Apple |
| Orange | Orange |
| Apple | Apple |
| Orange | Apple |
| Grapes | Grapes |
| Grapes | Grapes |
| Orange | Orange |
| Orange | Orange |
| Grapes | Apple |
| Grapes | Grapes |
| Apple | Orange |
| Orange | Orange |
| Apple | Apple |

*Hint: You can read about micro and macro averaging here:*

- *https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html*
- *https://www.evidentlyai.com/classification-metrics/multi-class-metrics*

1. **Task (9 Points)** Fill in the confusion matrix.

| | | Predicted | | |
|---|---|---|---|---|
| | | **apple** | **grapes** | **oranges** |
| Actual | **apple** | | | |
| | **grapes** | | | |
| | **oranges** | | | |

2. **Task (9 Points)** Fill in the precision, recall, and $F_1$-score per class.

| | | Metrics | | |
|---|---|---|---|---|
| | | Precision | Recall | $F_1$ |
| Actual | **apple** | | | |
| | **grapes** | | | |
| | **oranges** | | | |

3. **Task (5 Points)** What's the overall $F_1$-score for the classifier with *micro-averaging*?
4. **Task (5 Points)** What's the overall $F_1$-score for the classifier with *macro-averaging*?
5. **Task (3 Points)** When would the results of averaging metrics differ more?