The Great Computer Challenge, 2025
Artificial Intelligence (A.I.), Level 4
Problem 1

Background

The study of plant species is a fundamental aspect of botany and environmental science. Accurate classification of plant species based on their physical characteristics is essential for biodiversity research, conservation efforts, and ecological studies. The Iris Flower Dataset is a widely used benchmark in machine learning and statistics, providing a structured way to explore classification techniques. This problem challenges participants to analyze the Iris dataset, identify key distinguishing features, and develop a classification model to classify Iris flowers into their respective species.

Guidelines & Requirements

Participants will work with the Iris Flower Dataset, which consists of 150 samples, each representing an Iris flower. Each sample includes four numerical features: **sepal length**, **sepal width**, **petal length**, and **petal width**. Each sample belongs to one of three species: **Iris Setosa, Iris Versicolor, or Iris Virginica**.
Participants' tasks include exploring this dataset, developing a classification model, evaluating its performance, and gaining insights into the features that distinguish the different Iris species. The primary objectives are:

1. To analyze the dataset to identify the key features that differentiate the Iris species.
2. To develop a classifier to categorize an Iris flower into one of the three species based on its features.

Challenge 1

1. Conduct **exploratory data analysis** to understand the statistical distributions of the features and their relationships.

Challenge 2

2. Build **a classifier** on the given dataset using decision trees, support vector machines, k-nearest neighbors, logistic regression, neural networks, or other suitable methods.

- Split the dataset into training and testing sets (e.g., 80% training, 20% testing).
- Train the classifier on the training set and evaluate its performance on the testing set

Challenge 3

3. Evaluate model performance using the protocol and metrics specified in the Evaluation section.

Judging Criteria

The exploratory data analysis judging will focus on the ability to generate insightful visualizations and statistics:
(1) Highlight the dataset's characteristics (10 points) and
(2) The relationships between pairs of features (10 points).

The classification model will be evaluated based on the following protocol and metrics:
(1) Proper dataset splitting for training and testing (10 points),
(2) Performance evaluation using classification metrics such as accuracy, precision, recall, and F1-score (20 points), and
(3) Key features that influence classification and their real-world implications (10 points).

The judging criteria below override the default judging criteria.

Participating teams are allowed to use multiple computers to distribute the workload. Participants are not allowed to use any forms of large language models such as GPT, Gemini, and Llama during the competition.

The codes must be runnable on Google Colab and shared with fanchyna@gmail.com. The coding will be reviewed based on its design (5 points), smell (e.g., comments are clear and useful, good naming, etc.) (5 points), modularity (5 points), and reconfigurability (5 points).

The report, which is a single PDF file, should be sent to fanchyna@gmail.com. The report should describe the model(s), external data (if any), results (e.g., text, tables, figures, visualization, screenshots), the process to reproduce the results, discussion (if any), conclusion, references (if any). No page limit. No changes can be made after the competition ends. The report will be scored based on the model's performance (metrics in the evaluation section above), comprehensiveness, clarity, and accuracy. (report quality overall score: 20 points)

The judging committee reserves the right of post-competition validation. Awards will be revoked if plagiarism or fabrications are found.

A team's final score is determined by the consensus of all judges.

SOL Correlation

Not applicable.

Have fun and thanks for participating in the Great Computer Challenge, 2025!