中國醫藥大學附設醫院 1JZ0-醫研人工智慧演算法開發工程師上機試題

(1) Signal test questions

Predict whether the cancer is benign or malignant

Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

Attribute Information:

1. ID number
2. Diagnosis (M = malignant, B = benign)
3. Ten real-valued features are computed for each cell nucleus:
   - ·radius (mean of distances from center to points on the perimeter)
   - ·texture (standard deviation of gray-scale values)
   - ·perimeter
   - ·area
   - ·smoothness (local variation in radius lengths)
   - ·compactness ($perimeter2/area-1.0$)
   - ·concavity (severity of concave portions of the contour)
   - ·concave points (number of concave portions of the contour)
   - ·symmetry
   - ·fractal dimension ("coastline approximation" - 1)

The mean, standard error and "worst" or largest (mean of the three largest values) of these features were computed for each image,
resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius.

Missing attribute values: none

Class distribution: 357 benign, 212 malignant

Questions:

1. Show the correlation among all features.
2. Use one of the feature selection methods to select the 5 most important features.
3. Predict the risks by any of the ML models (SVM, logistic regression, random forests, xgboost, NN, etc.) using the features selected in 2. with the following data splitting:

   x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.25, random_state = 16)
4. Tune at least two of the hyperparameters (grid search, random search, ...) to improve the performance of the model.
5. Visualize the model performance (confusion matrix, precision-recall curve, F1-score, ...)
6. Please submit the code to produce the testing results.

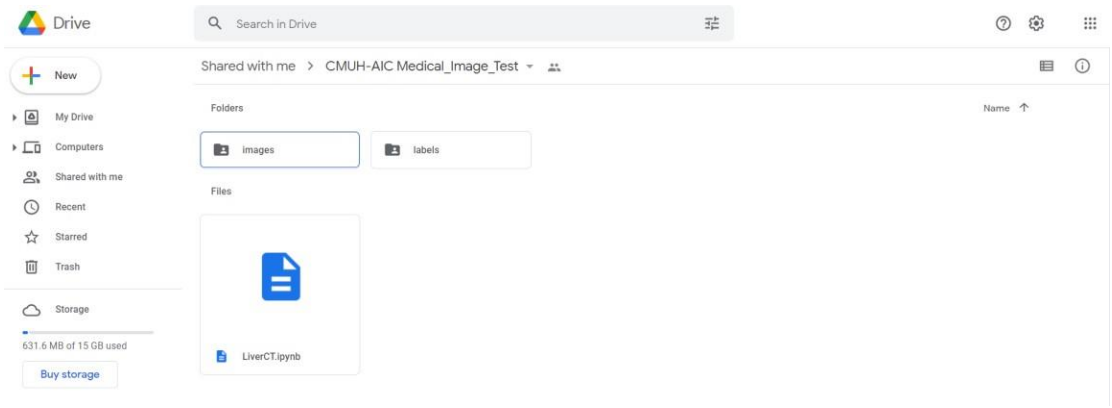
(2) Image test questions

中國醫藥大學附設醫院 人工智慧醫療診斷中心醫學影像題目

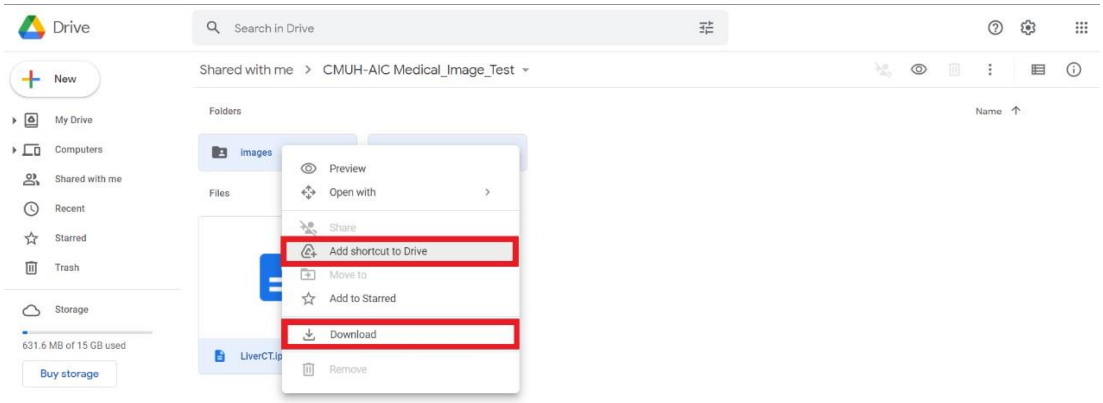
https://drive.google.com/drive/folders/1cXzaqbw9RhU5PB5lHQWQ11hnPIZNEXQF?usp=sharing

請先至上列網址將影像檔及標註檔下載或是加入自己的 Google Drive 中，使用方法如下。

1. 進入下載檔案網址



2. 全選檔案選擇下載或是複製至自己的雲端硬碟中。



3. 請 在 Test.ipynb 內作答，並回傳此檔案。若有其他補充資料可另以其他文字檔或圖檔一併附上。

(3) NLP test questions

Text Classification on IMDb

Dataset information:
http://ai.stanford.edu/~amaas/data/sentiment/

Dataset URL:
http://ai.stanford.edu/~amaas/data/sentiment/aclImdb_v1.tar.gz

Questions:

1. Train a model based on any machine learning method.

2. Visualize the performance of the model (confusion matrix, precision-recall curve, F1-score, ...).

3. Please briefly explain what you do in each training step.

4. Please submit your model and the codes to produce the testing results.