

HW2 說明

請使用 GDSC cell line gene expression 來建立(機器學習)模型，並預測 drug 對 testset(CCLE)的 response(resistant, sensitive)。

資料夾包含 3 個檔案(GDSC_PDX_Paclitaxel.csv, CCLE_PDX_Paclitaxel.csv 和 GDSC_Paclitaxel_info.csv)。

請使用 GDSC_PDX_Paclitaxel.csv 和 GDSC_Paclitaxel_info.csv 2 個資料集 cell lines 名字相同當 training set。

GDSC_PDX_Paclitaxel 檔案包含 399 個 GDSC training cell lines gene expression level。

CCLE_PDX_Paclitaxel 檔案包含 469 個 CCLE test cell lines gene expression level。

GDSC_PDX_Paclitaxel_info 檔案欄位說明如下:

IC50: kill 50% cancer 細胞的藥物濃度。IC50 的值越小，表示藥物以很小的劑量就可以抑制 50% 細胞生長，也就代表細胞株對藥物是敏感的(sensitive (S))；而 IC50 的值越大，就代表細胞株對於藥物是有抗藥性的 (resistant (R))。

MAX_CONC_MICROMOLAR: 藥物對細胞株使用的最大用藥濃度 (maximal tested concentration)，若 IC50 小於最大用藥濃度，則標記為 S；若 IC50 大於最大用藥濃度，則標記為 R。