Homework 3
Due: Midnight, April 6st, 2022

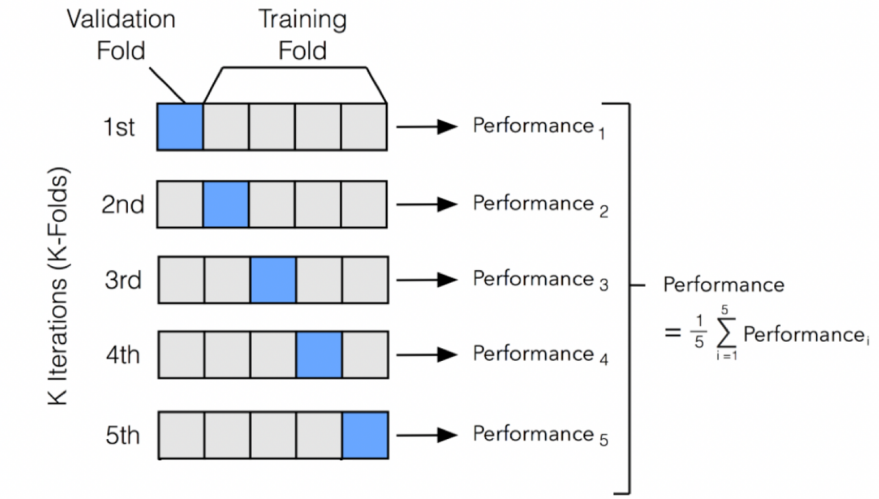
Description:
We will compare linear model, lasso, Elastic Net, SVM, and Naive Bayes on predicting drug properties.

Write everything in a jupyter notebook. Write your code in the "code" section in jupyter notebook and your answers to the questions in the "Markdown" section. You can use python package to do the following tasks.

1. Data preparation (10 points):
1.1 Install deepchem: https://deepchem.readthedocs.io/en/latest/get_started/installation.html
1.2 Load BBBP Datasets and convert SMILES to figure prints.
https://deepchem.readthedocs.io/en/latest/api_reference/moleculenet.html#bbbp-datasets

"name" - Name of the compound
"smiles" - SMILES representation of the molecular structure
"p_np" - Binary labels for penetration/non-penetration

2. Evaluation (40 points)
2.1 For binary classification, use AUPR score to evaluate the classification performance. (Hint: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html)
2.2 For each method, using the 5 iteration 5-fold cross validation as the final metric for each method. Summarize the results in a table.



2. Linear Model (10 points).
3. Lasso (10 points).
4. Elastic Net (10 points).

5. SVM (10 points).
6. Naïve Bayes (10 points).