NYCU Pattern Recognition, Homework 4

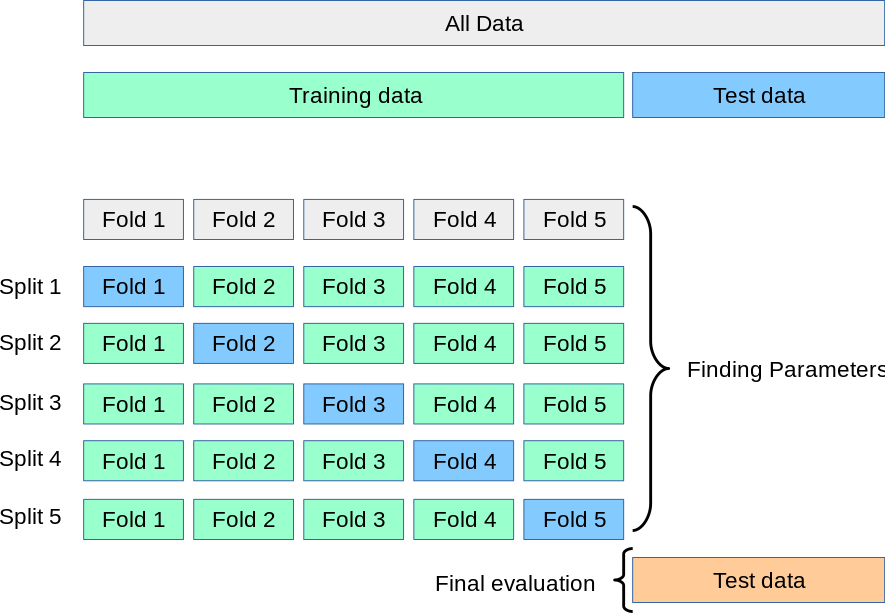
**Deadline: May 17, 23:59**

**Part. 1, Coding (50%)**:

For this coding assignment, you are required to implement Cross-Validation and Grid Search using only NumPy. After that, you should train the SVM model from scikit-learn on the provided dataset and test the performance with the testing data. **You will get no points by simply calling** [**sklearn.model\_selection.GridSearchCV**](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)**.**

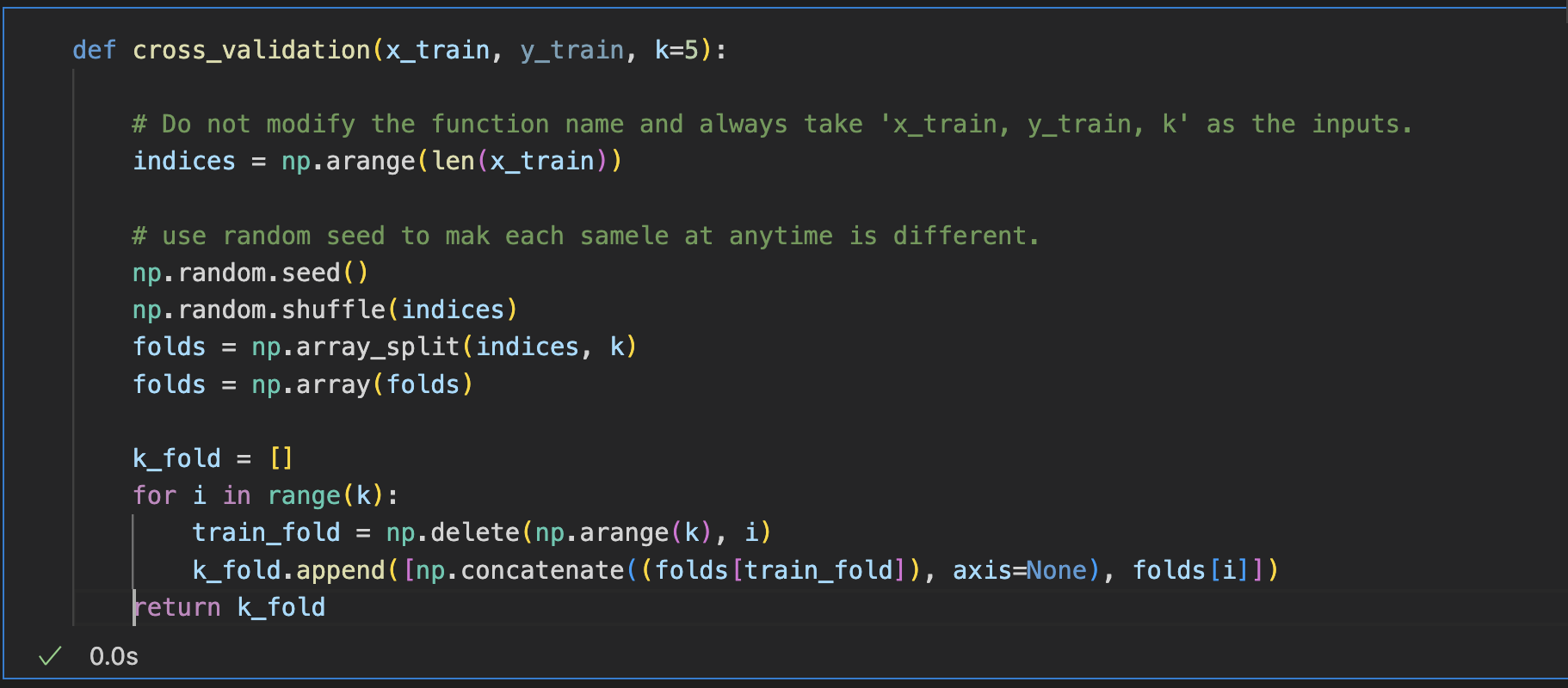
**(50%) K-Fold Cross-Validation & Grid Search**

**Requirements:**

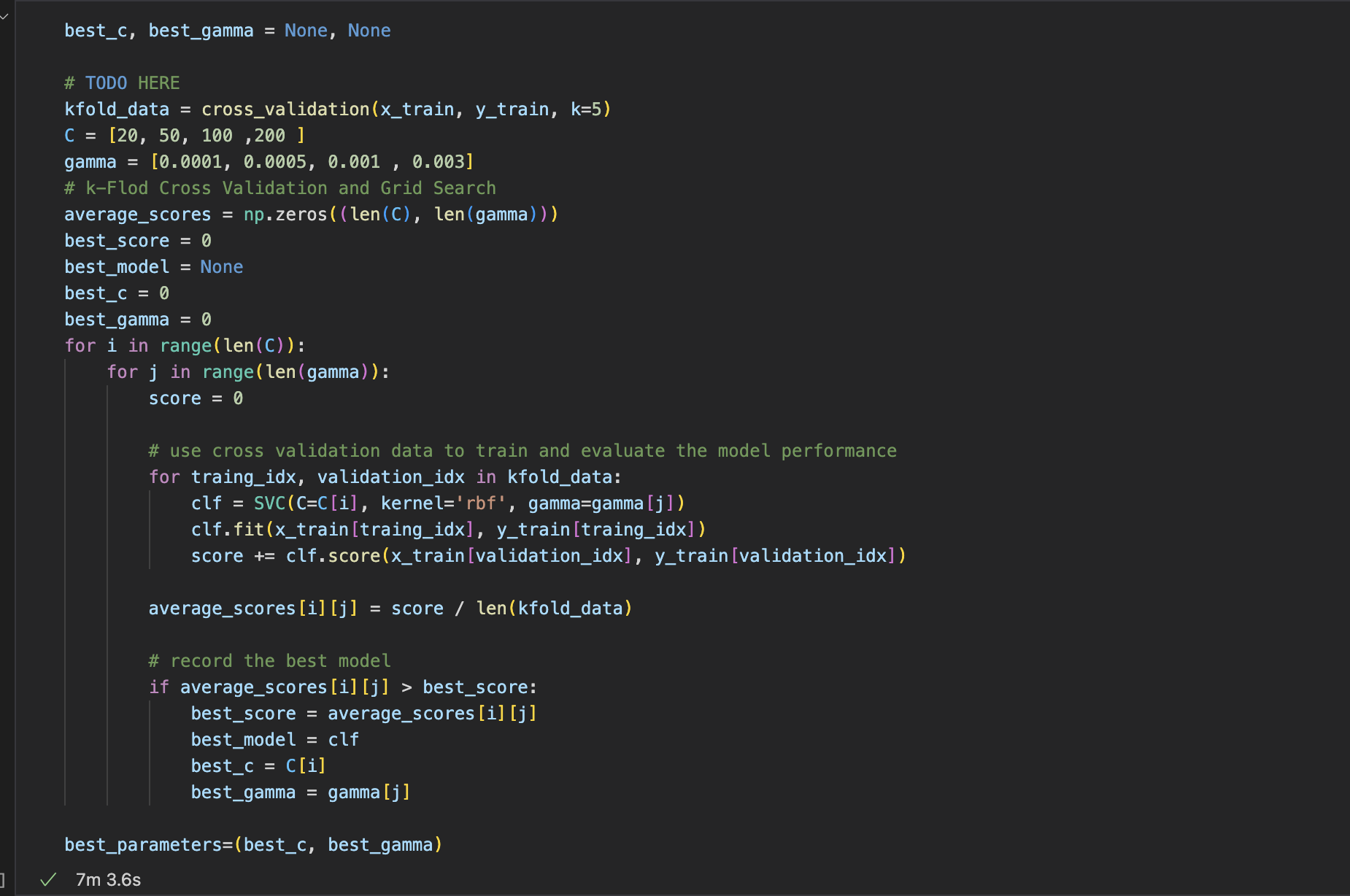
* Implement **K-Fold Cross-Validation** by creating a function that takes K as an argument and returns a list of K sublists.
  + Each sublist should contain two parts:
    - The first part contains the index of all training folds (index\_x\_train, index\_y\_train), for example, Fold 2 to Fold 5 in split 1.
    - The second part contains the index of the validation fold (index\_x\_val, index\_y\_val), for example, Fold 1 in split 1 .
  + You need to handle if the sample size is not divisible by K.
  + The first **n\_samples % n\_splits** folds should have a size of **n\_samples // n\_splits + 1**, and the other folds should have a size of **n\_samples // n\_splits**. Here, n\_samples is the number of samples and n\_splits is K.
  + Each of the samples should be used **exactly once** as the validation data.
  + Please **shuffle** your data before partition.
* Implement **Grid Search & Cross-Validation**:
  + Using [sklearn.svm.SVC](https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html) to train a classifier on the provided train set and perform **Grid Search** to find the best hyperparameters via cross-validation.

**Criteria:**

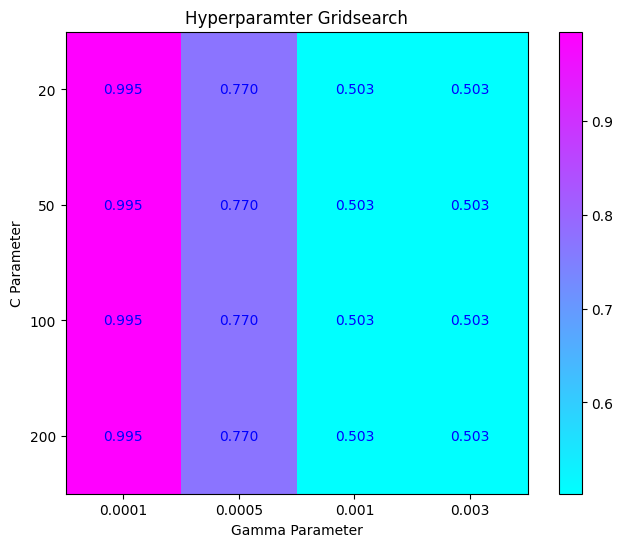
1. (10%) Implement K-fold data partitioning.



1. (10%) Set the kernel parameter to 'rbf' and do grid search on the hyperparameters **C** and **gamma** to find the best values through cross-validation. Print the best hyperparameters you found. Note that we suggest using K=5 for the cross-validation.

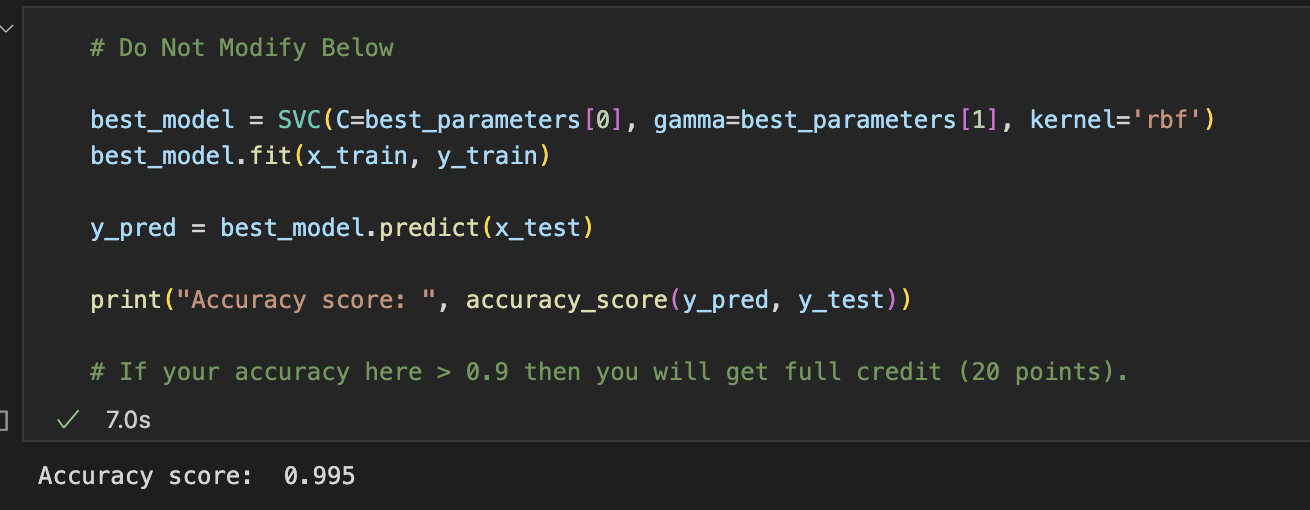


1. (10%) Plot the results of your SVM's grid search. Use "gamma" and "C" as the x and y axes, respectively, and represent the average validation score with color. Below image is just for reference.



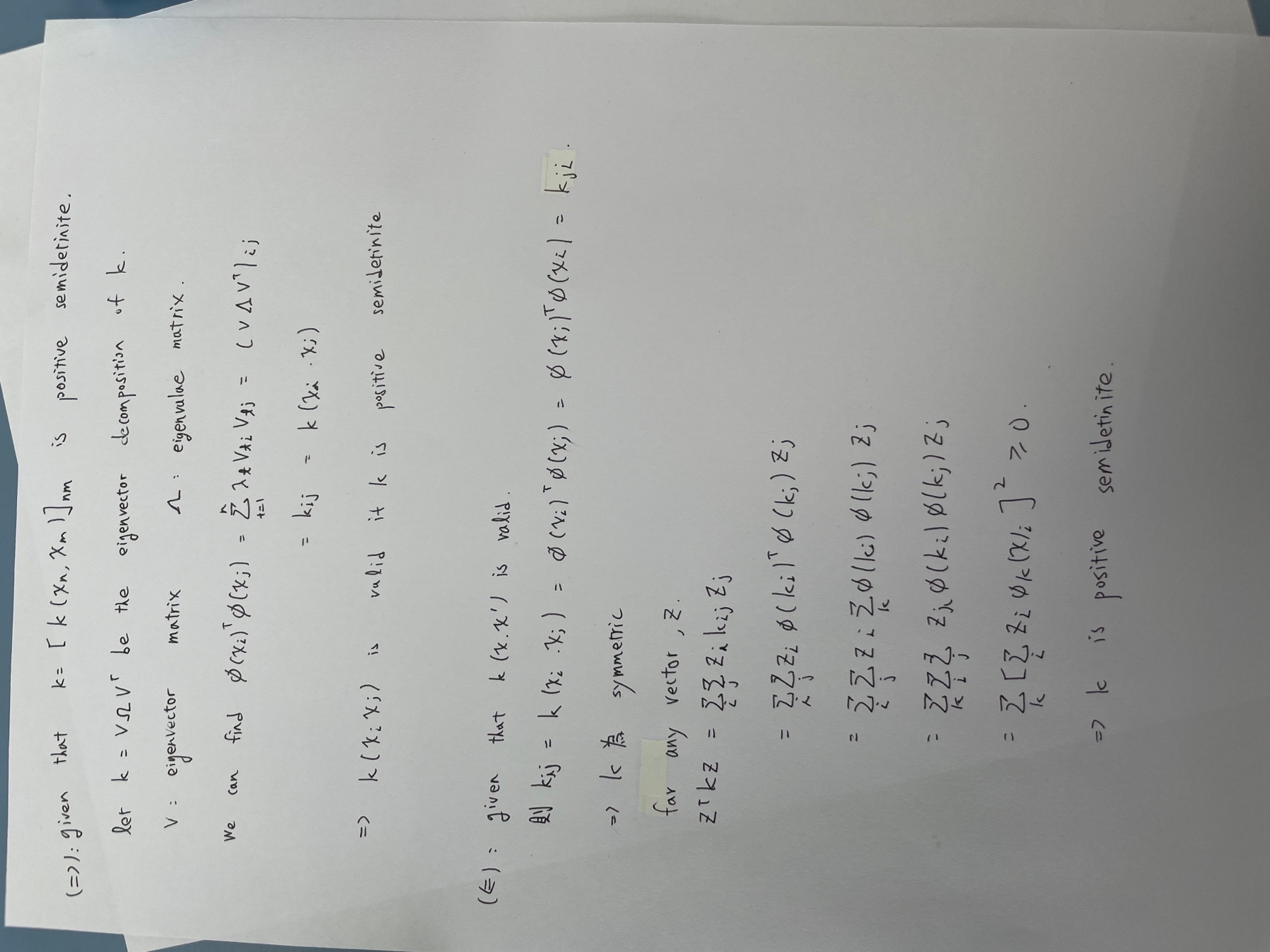
1. (20%) Train your SVM model using the best hyperparameters found in Q2 on the entire training dataset, then evaluate its performance on the test set. Print your testing accuracy.

|  |  |
| --- | --- |
| **Points** | **Testing Accuracy** |
| **20 points** | **acc > 0.9** |
| **10 points** | **0.85 <= acc <= 0.9** |
| **0 points** | **acc < 0.85** |

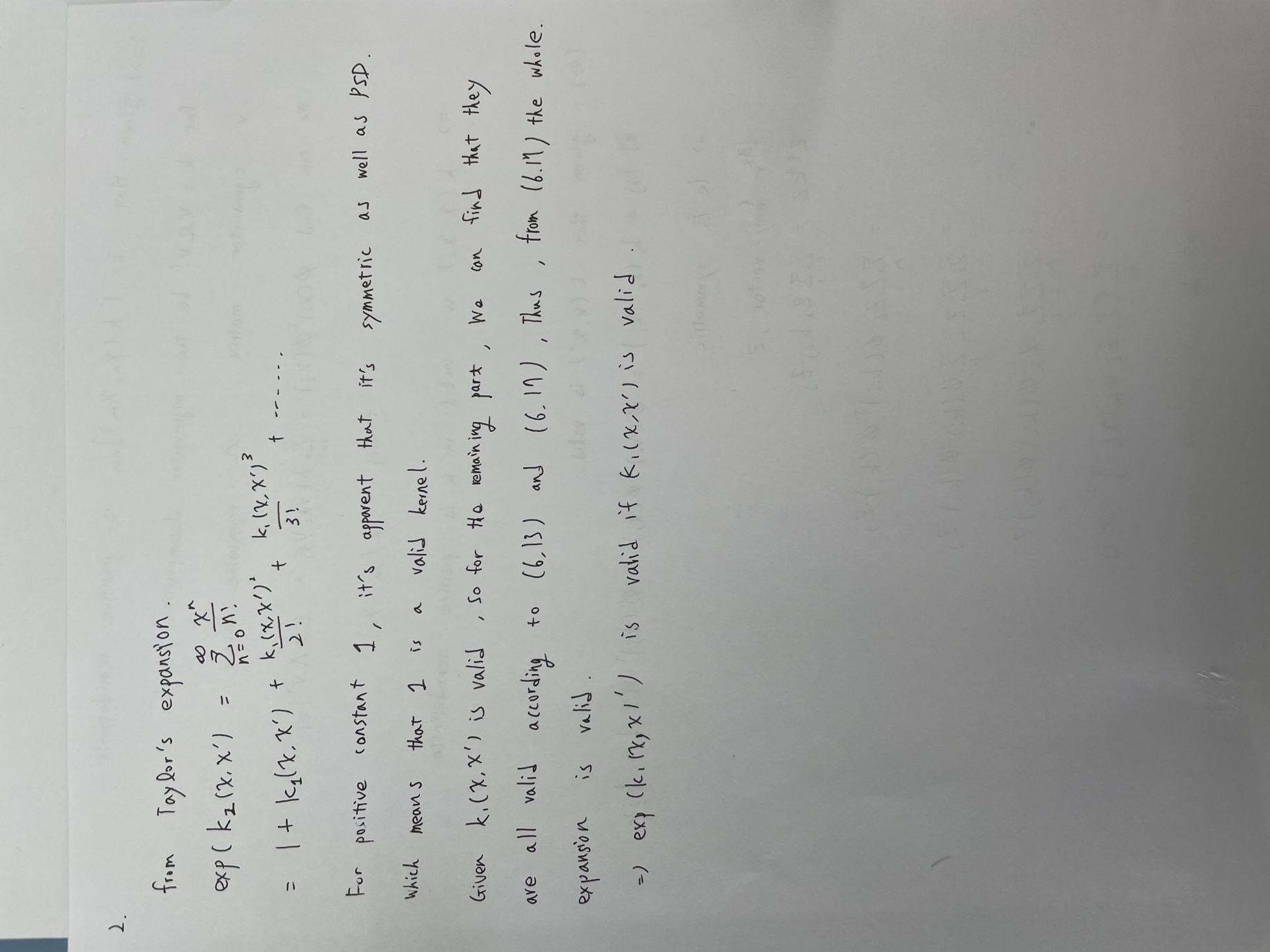
****

**Part. 2, Questions (50%):**

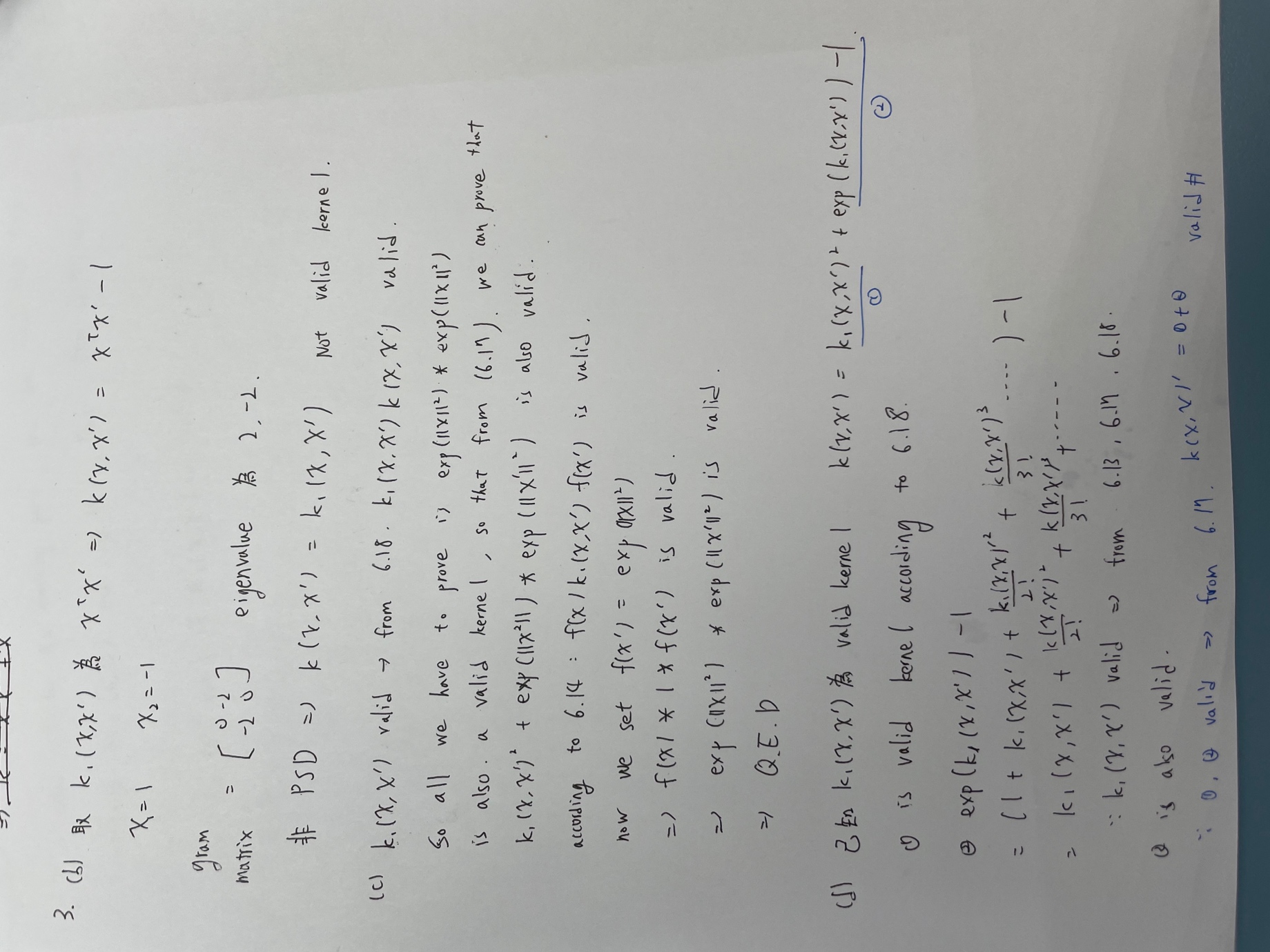
1. (10%) Show that the kernel matrix should be positive semidefinite is the necessary and sufficient condition for to be a valid kernel.



1. (10%) Given a valid kernel , explain that is also a valid kernel. (Hint: Your answer may mention some terms like \_\_\_\_ series or \_\_\_\_ expansion.)



1. (20%) Given a valid kernel , prove that the following proposed functions are or are not valid kernels. If one is not a valid kernel, give an example of that the corresponding is not positive semidefinite and show its eigenvalues.

**

1. Consider the optimization problem

State the dual problem. (Full points by completing the following equations)

= \_\_\_\_\_\_\_\_\_(x - 2)^2 + λ[(x + 4)(x - 1) – 3]\_\_\_\_\_\_\_\_\_\_\_\_

= \_\_\_\_\_\_\_\_\_\_\_2(x - 2) + λ(2x + 3) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

when ,

= \_\_\_\_\_\_\_\_\_(4 - 3λ)/(2 + 2λ) \_\_\_\_\_\_\_\_\_\_\_

= [(4 - 3λ) / (2 + 2λ) - 2]^2 + λ{[(4 - 3λ) / (2 + 2λ)) + 4][((4 - 3λ) / (2 + 2λ)) – 1]– 3}