NYCU Introduction to Machine Learning, Homework 4

110550014, 吳權祐

The screenshot and the figures we provided below are just examples. The results below are not guaranteed to be correct. Please make sure your answers are clear and readable, or no points will be given. Please also remember to convert it to a pdf file before submission. You should use English to answer the questions. After reading this paragraph, you can delete this paragraph.

Part. 1, Coding (50%):

1. (10%) Show the accuracy score of the testing data using *linear_kernel*. Your accuracy score should be higher than 0.8.

```
Accuracy of using linear kernel (C = 0.1):  0.83
```

2. (20%) Tune the hyperparameters of the *polynomial_kernel*. Show the accuracy score of the testing data using *polynomial_kernel* and the hyperparameters you used.

```
Accuracy of using polynomial kernel (C = 0.1, degree = 4):  0.99
```

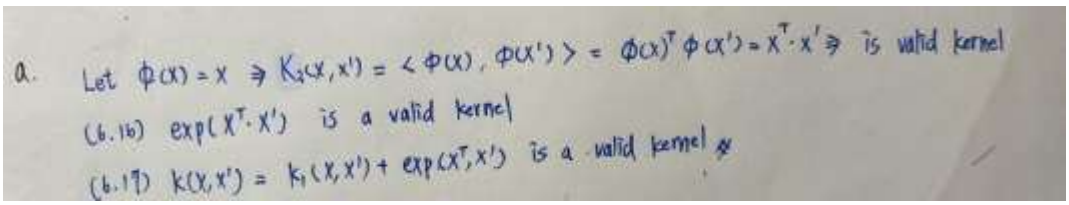
3. (20%) Tune the hyperparameters of the *rbf_kernel*. Show the accuracy score of the testing data using *rbf_kernel* and the hyperparameters you used.

```
Accuracy of using rbf kernel (C = 1.5, gamma = 0.5):  0.98
```

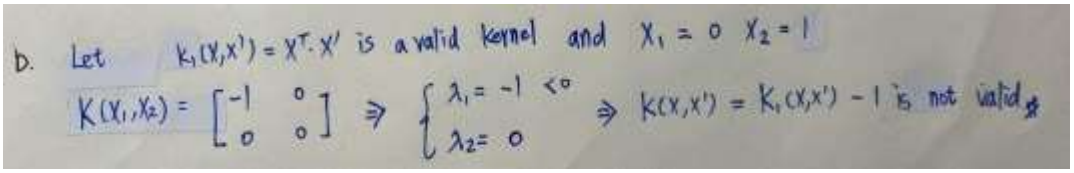
Part. 2, Questions (50%):

1. (20%) Given a valid kernel $k_1(x, x')$, prove that the following proposed functions are or are not valid kernels. If one is not a valid kernel, give an example of $k(x, x')$ that the corresponding K is not positive semidefinite and shows its eigenvalues.

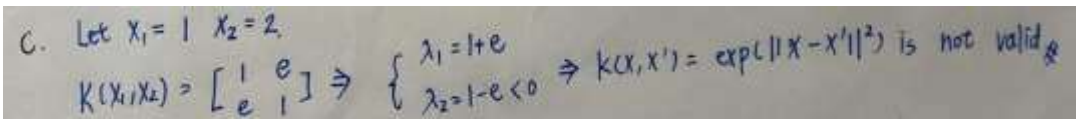
   a. $k(x, x') = k_1(x, x') + exp(x^T x')$



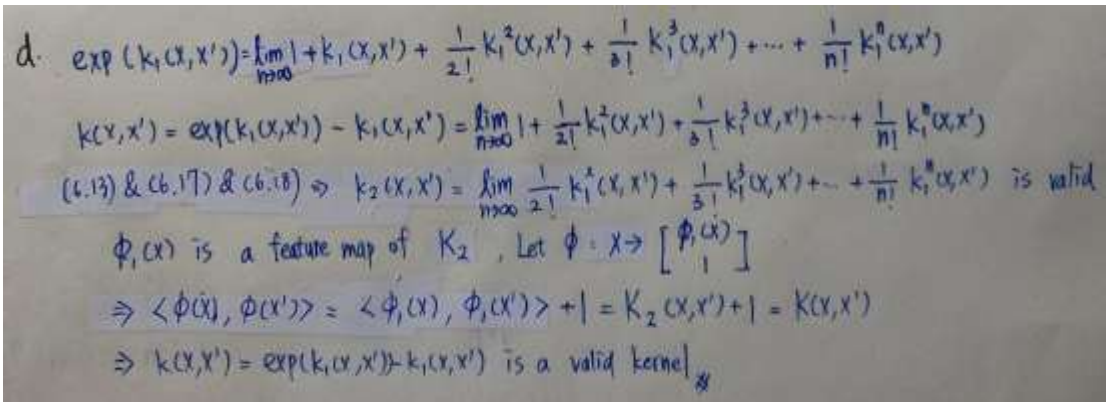
   b. $k(x, x') = k_1(x, x') - 1$



   c. $k(x, x') = exp\left(\|x - x'\|^2\right)$



d. $k(x, x') = exp\left(k_1(x, x')\right) - k_1(x, x')$



2. (15%) One way to construct kernels is to build them from simpler ones. Given three possible "construction rules" : assuming $K_1(x, x')$ and $K_2(x, x')$ are kernels then so are

   a. (scaling) $f(x)K_1(x, x')f(x'), f(x) \in R$

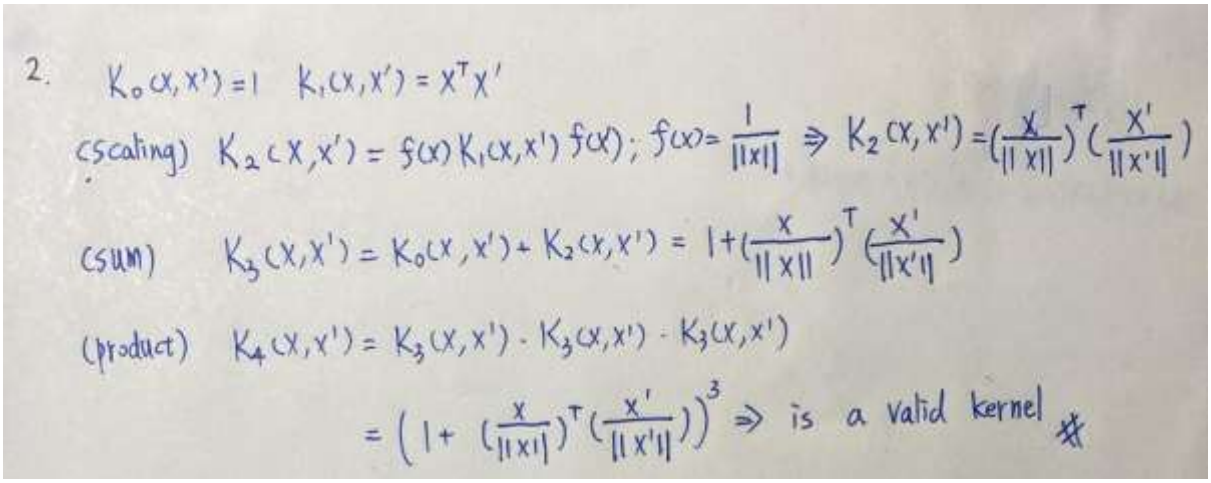
   b. (sum) $K_1(x, x') + K_2(x, x')$

   c. (product) $K_1(x, x')K_2(x, x')$

Use the construction rules to build a normalized cubic polynomial kernel:

$$K(x, x') = \left(1 + \left(\frac{x}{||x||}\right)^T \left(\frac{x'}{||x'||}\right)\right)^3$$

You can assume that you already have a constant kernel $K_0(x, x') = 1$ and a linear kernel $K_1(x, x') = x^T x'$. Identify which rules you are employing at each step.



3. (15%) A social media platform has posts with text and images spanning multiple topics lik e news, entertainment, tech, etc. They want to categorize posts into these topics using SV Ms. Discuss two multi-class SVM formulations: `One-versus-one` and `One-versus-the-res t` for this task.

I use OvO tp represent One-versus-one and OvR to represent One-versus-the rest in the following questions.

a. The formulation of the method [how many classifiers are required]

   If there are n classes need to be classified, in OvO formulation, we need to train n(n-1) / 2 classifiers, and in OvR, there are only n classifiers need to be trained.

b. Key trade offs involved (such as complexity and robustness).

   Because of the number of classifiers need to be trained, OvO has higher complexity on training. But since the classifiers in OvO could distinguish two specific classes better than classifiers in OvR, OvO should be more robust.

c. If the platform has limited computing resources for the application in the inference phase and requires a faster method for the service, which method is better.

   If there are some computing resources limitation, and require a faster method for the service, OvR is a better method, since it has lower cost for training, and lower number of classifiers. This could have faster process time, when it needs to do classifying.