# BDA Lab3 - Machine Learning

Student: Xuan Wang (xuawa284)
Student: Lepeng Zhang (lepzh903)

See the code lab3.py

**Question1:**

SUM kernel prediction results:

```
[('24:00:00', 5.8), ('22:00:00', 5.7), ('20:00:00', 5.6), ('18:00:00', 5.7), ('16:00:00', 5.9),
('14:00:00', 6.0), ('12:00:00', 6.1), ('10:00:00', 5.9), ('08:00:00', 5.6), ('06:00:00', 5.2),
('04:00:00', 5.3)]
```

The choice of width significantly influences the estimator's performance for the Gaussian kernel. If the bandwidth is too small, the estimator may be too sensitive to small fluctuations in the data (overfitting). If it's too large, important details may be smoothed out (underfitting). In this case we chose h_distance = 500, h_date = 10, h_time = 2 , which is made through trial and error. Each of these parameters also meet the assumptions that time might has a significant impact on the temperature, thus a smaller h_time is used. While, distance may have less influence on the temperature, thus a larger h_distance is used.

**Question 2:**

PROD kernel prediction results:

```
[('24:00:00', 9.2), ('22:00:00', 9.9), ('20:00:00', 11.9), ('18:00:00', 14.4), ('16:00:00', 16.0),
('14:00:00', 16.8), ('12:00:00', 16.7), ('10:00:00', 15.5), ('08:00:00', 13.5), ('06:00:00', 11.1),
('04:00:00', 8.7)]
```

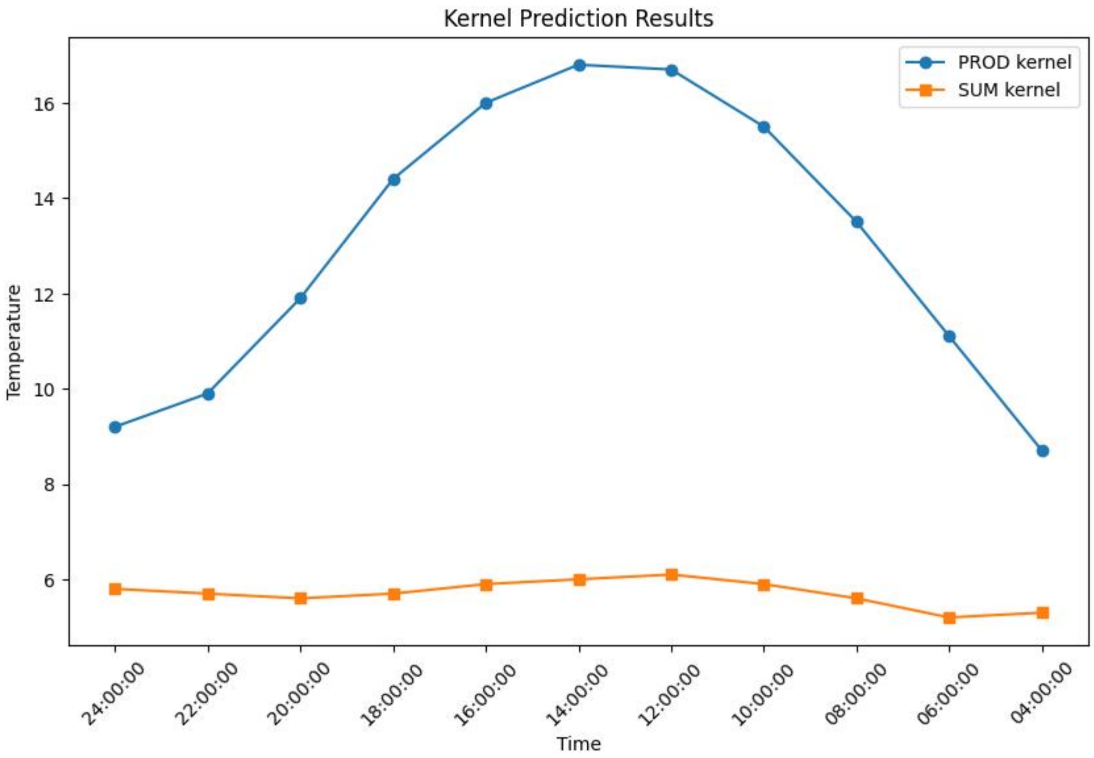

Figure 1

Normally when using sum kernels, it's essentially giving equal importance to all three factors (distance, date, and time) in the model. This means that even if one factor has a strong indication (like a very close time), if the other factors are not as indicative, the final result may still not be very high. This can sometimes lead to underfitting, where the model does not capture all the patterns in the data.

On the other hand, when taking product of kernels, it's multiplying the results together. This means that if any one of the factors has a strong indication (a high kernel value), it can have a significant impact on the final result, even if the other factors are not as indicative. This can sometimes lead to overfitting, where the model captures not only the underlying pattern but also the noise in the data.

In our case, according the Figure 1, when using the product of kernels, the predicted temperature are relatively higher than that using the sum of kernels, and the product of kernels is giving more accurate results with a diurnal temperature variation. This could be because the factors in our model (distance, date, and time) are not equally important, and the product of kernels is able to capture the interactions between these factors more effectively than the sum of kernels.