HMW1 Pratical report

fahed.elourajini

September 2020

1 Question 5:

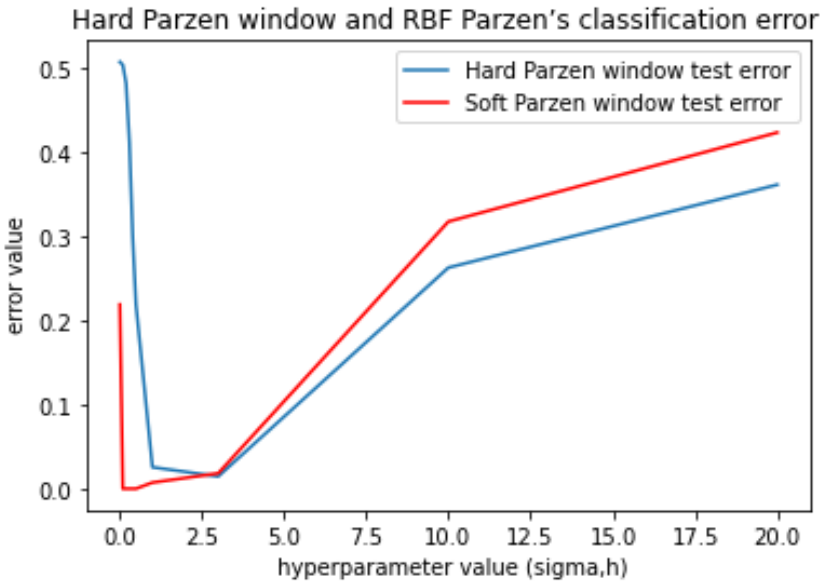

Figure 1: Hard parzen window and RBF paren's classification error.

As we see in this plot, with near 0 values '0.01' for the hyper parameters $\sigma$and **h** the error values is huge for both algorithms 'Soft and Hard parzen windows' however by increasing the value of:

σ: the error rate of the soft parzen algorithm will decrease, while σ values is under 0.2. When the σ values pas the 0.2 value the error start increasing step by step until the value of 3.0. when the sigma reach the 3.0 value , the error start increasing exponentially with the increasing of the σ parameter value

h: the error rate of the hard parzen algorithm will decrease, the error for hard parzen algorithm start decreasing when the h parameters is between the 0.1 and the 3.0 values, but when the value of h become more the 3.0 the error of hard parzen start increasing exponentially.

PS:
we see that betwwen the values 0.1 and 3.0 whenever we increase the values of h and σ the error of soft parzen algorithm is lesser then the hard parzen window algorithm however starting from 3.0, whenever we increase the σ or the h parameters the error for hard parzen will still less then the error of soft parzen algorithm.
As a conclusion if the h and σ parameters are too big then the error value will be big , also if thy are too small exmpl'0.01' the error also will be big=¿ the best value are between 0.1 and 3.0. The best value for h is 3.0 and the best value for σ is 0.2

```
import matplotlib.pyplot as plt
sigma=[0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 1.0, 3.0, 10.0, 20.0]
h=[0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 1.0, 3.0, 10.0, 20.0]
train,valid,test=split_dataset(banknote)
e=ErrorRate(train[:,:-1],train[:,-1],valid[:,:-1],valid[:,-1])
plt.plot(h, [e.hard_parzen(hi) for hi in h], label='Hard Parzen window test error')
plt.plot(h, [e.soft_parzen(sigmai) for sigmai in sigma], color='red',label='Soft Parzen window test error')
plt.title("Hard Parzen window and RBF Parzen's classification error ")
plt.legend()
plt.xlabel(' hyperparameter value (sigma,h)')
plt.ylabel(' error value')
plt.show()
```

Figure 2: Hard parzen window and RBF paren's classification error Code.

2 Question 7:

the running time complexity of the soft and hard parzen algorithms:

- **Hard parzen window**: Hard parzen window is a kind of kNN technique a specific number of data points will participate in the vote=¿ only the data point that fall in a range of distance will participate in the vote.
  so the complexity time of the hard parzen window is :
  $O(d)$= to calculate the distance to one example

  $O(nd)$= to find the distance of the n examples

  $O(knd)$= to find the vote of the nearest neighbors that respect the distance condition
  => The complexity of the Hard parzen window is $O(knd)$

- **Soft parzen window**: Soft parzen window is a little bit similar to the Hard parzen technique, however instead of only a set of point that partic-

ipate in the vote, in soft parzen window all of the points will participate in the vote.
so the complexity time of the soft parzen window is :
$O(d)$= to calculate the distance to one example

$O(nd)$= to find the distance of the n examples

$O(n.nd) = O(n^2 \cdot d)$ = to find the total vote of all the data points
=> The complexity of the Hard parzen window is $O(n^2 \cdot d)$

3 Question 9:

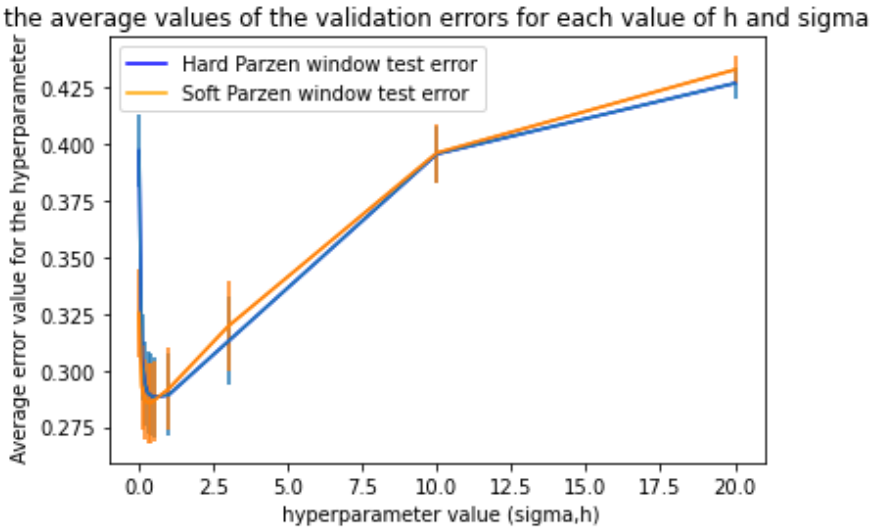

Figure 3: Hard parzen window and RBF paren's classification average error.

The difference between the error on the Q5 and the error on the Q9:

as we see in the figure the average errors between the two algorithms 'soft and hard parzen window' are close to each other comparing to the first plot in the Q5. Even when we increase the value of hyperparameters the difference betwenn the two erro is too small, However in the first plot of the Q5, whenever we increase the hyper parameters the distance between the two error line increases.
Also in the other plot we notecide that the error started increasing from the value 3.0 of hyper parameter h , however in this plot we see that the error

started increasing from the value of 1.0 for h and 0.25 for *sigma*

=> in order to make this obvious we will change the x and y axis lemitation :

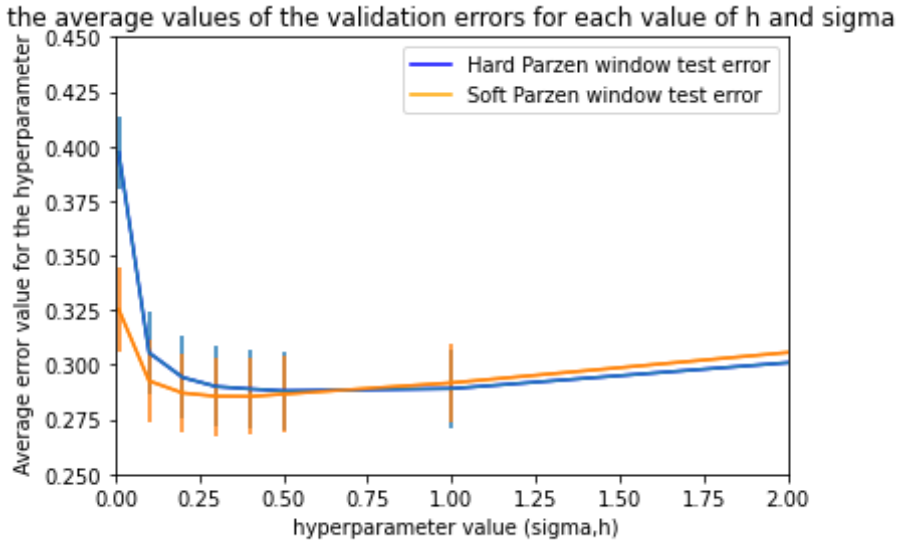

Figure 4: Hard parzen window and RBF paren's classification average error.

```
import matplotlib.pyplot as plt
sigma=[0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 1.0, 3.0, 10.0, 20.0]
h=[0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 1.0, 3.0, 10.0, 20.0]
plt.xlim(0,2)
plt.ylim(0.25,0.45)

plt.errorbar(h,har_parzen_error_projection_matrix.mean(axis=0),yerr=har_parzen_error_projection_matrix.std(axis=0)*0.2)
plt.plot(h,har_parzen_error_projection_matrix.mean(axis=0),color="blue",label='Hard Parzen window test error')
plt.errorbar(h,soft_parzen_error_projection_matrix.mean(axis=0),yerr=soft_parzen_error_projection_matrix.std(axis=0)*0.2)
plt.plot(h,soft_parzen_error_projection_matrix.mean(axis=0),color="orange",label='Soft Parzen window test error')

plt.title("the average values of the validation errors for each value of h and sigma")
plt.legend()
plt.xlabel(' hyperparameter value (sigma,h)')
plt.ylabel(' Average error value for the hyperparameter')
plt.show()
```

Figure 5: Hard parzen window and RBF paren's classification average error source code.