Name: Omer Sayem

Student ID: 40226505

Email: sayemomer@gmail.com

1. Kernels

a. A function K(x,z) to be the number of unique words that occur in both,

a)

Suppose,

$x =$ "I love Machine"

$z =$ "machine is not human"

Size of $D = 4$,

$D = \{$"I", "love", "Machine", "human"$\}$

kernel function k is a Dot product of feature map $\phi(x)$ such that ith word of D is 1 if exits or 0. So,

$$k(x,z) = \phi(x)^T \phi(z)$$

$$\phi(x)^T = [1,1,1,0]$$

$$\phi(z) = [0,0,1,0]$$

So, $k(x,z) = [1,1,1,0] \times [0,0,1,0]$

$$= (1.0) + (1,0) + (1.1) + (0.0)$$

$$k(x,z) = 1$$

$\therefore$ Exactly 1 match. "Machine".

function definition,

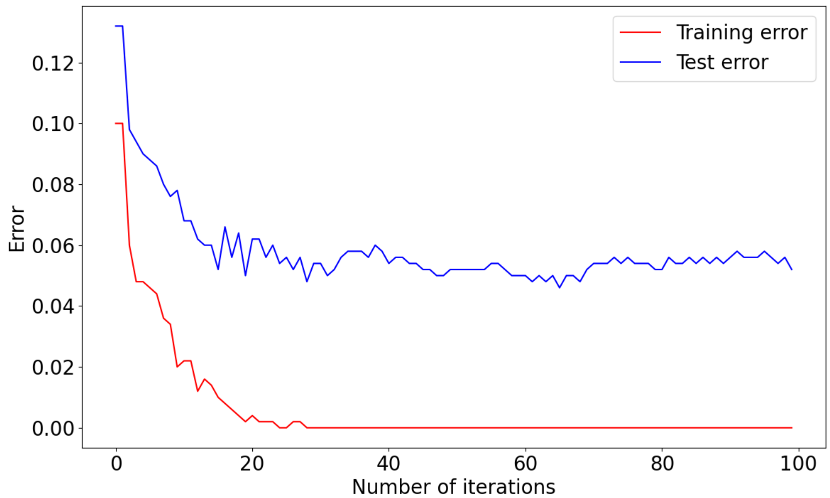
$$k(x,z) = \phi(x)^T . \phi(z) = \sum_{i=1}^{D} I_{w_i}(x)\, I_{w_i}(z)$$

$|\ I_w(t) =$ Indicator function $(0,1)$

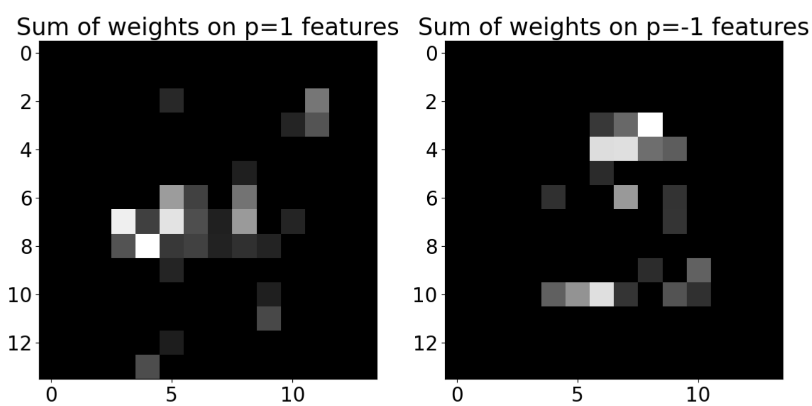
This function is a kernel function because it is written as an inner product ϕ(x)T ϕ(z) for feature mapping of the vocabulary D.

2. Programming: AdaBoost

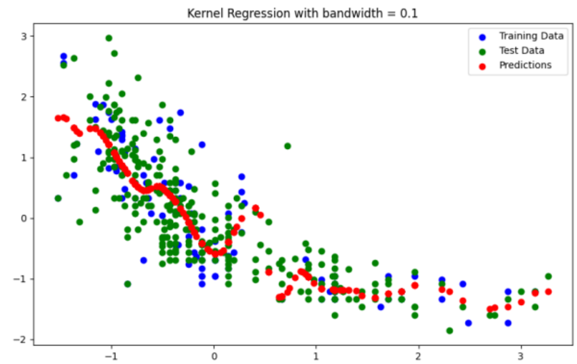
Plot of training error and test error using AdaBoost Algorithm,

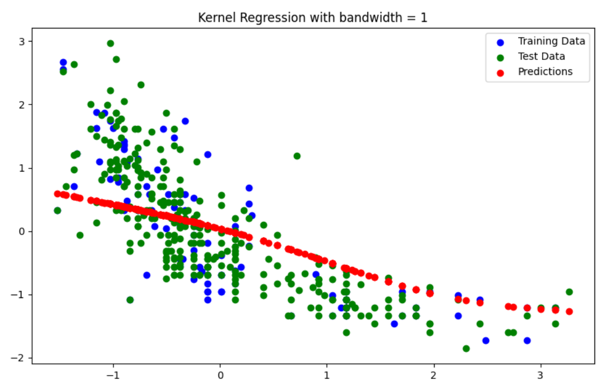


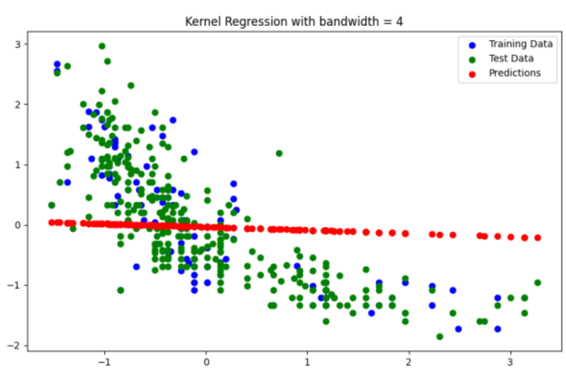
Visualization of the final classifier produced using visualizeClassifier in utils.py



3. Programming: Kernel regression



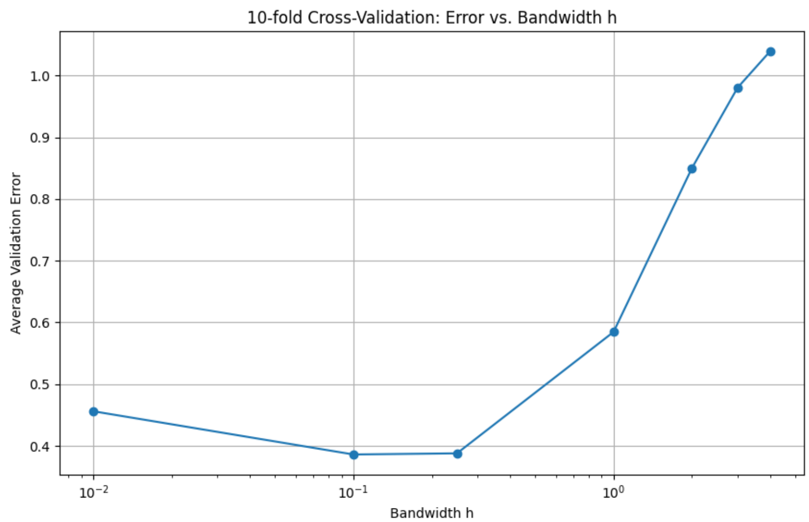




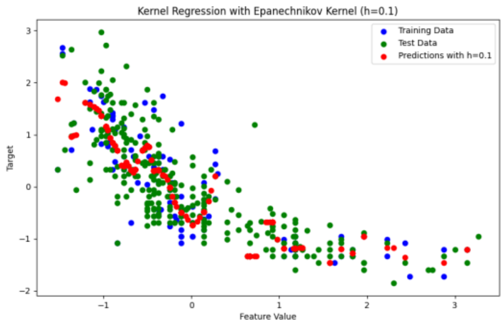
A) For small bandwidth h=0.1, we can see that the model fluctuates frequently which might lead to overfitting as it captures the noise from training data.

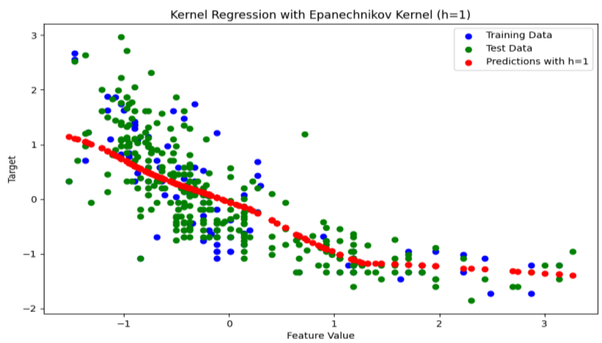
For moderate bandwidth h=1, the graph shows a general trend in the training data and is much smoother than the previous one.

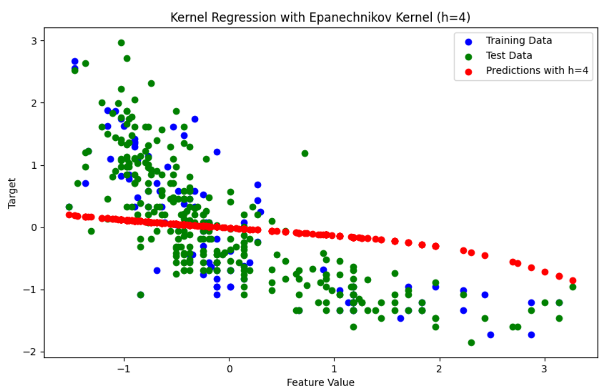
For High bandwidth h=4, the line is horizontal, which might miss some of the underlying data results in underfitting.



B) I would choose the bandwidth value h = 0.01 from the cross-validation error because it minimizes the average error.



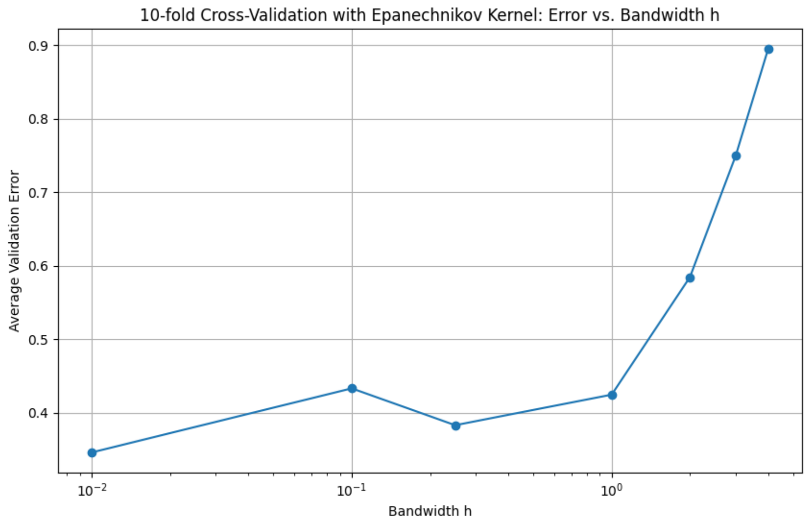




C) For small bandwidth h=0.1, we can see a similar trend to the Gaussian kernel, that the model fluctuates frequently which might lead to overfitting as it captures the noise from training data.

For moderate bandwidth h=1, comparatively, the Gaussian kernel was a bit smoother, but here the graph shows a general trend in the training data and is much smoother than the previous one.

For High bandwidth h=4, the line is horizontal indicating high bias like the Gaussian Kernel, which might miss some of the underlying data results in underfitting.



D) This curve differs a bit from the Gaussian, but still the optimal is around 0.1 to 1 where the error reaches its minimum. Thus, I would choose the bandwidth value h = 0.01 from the cross-validation error.