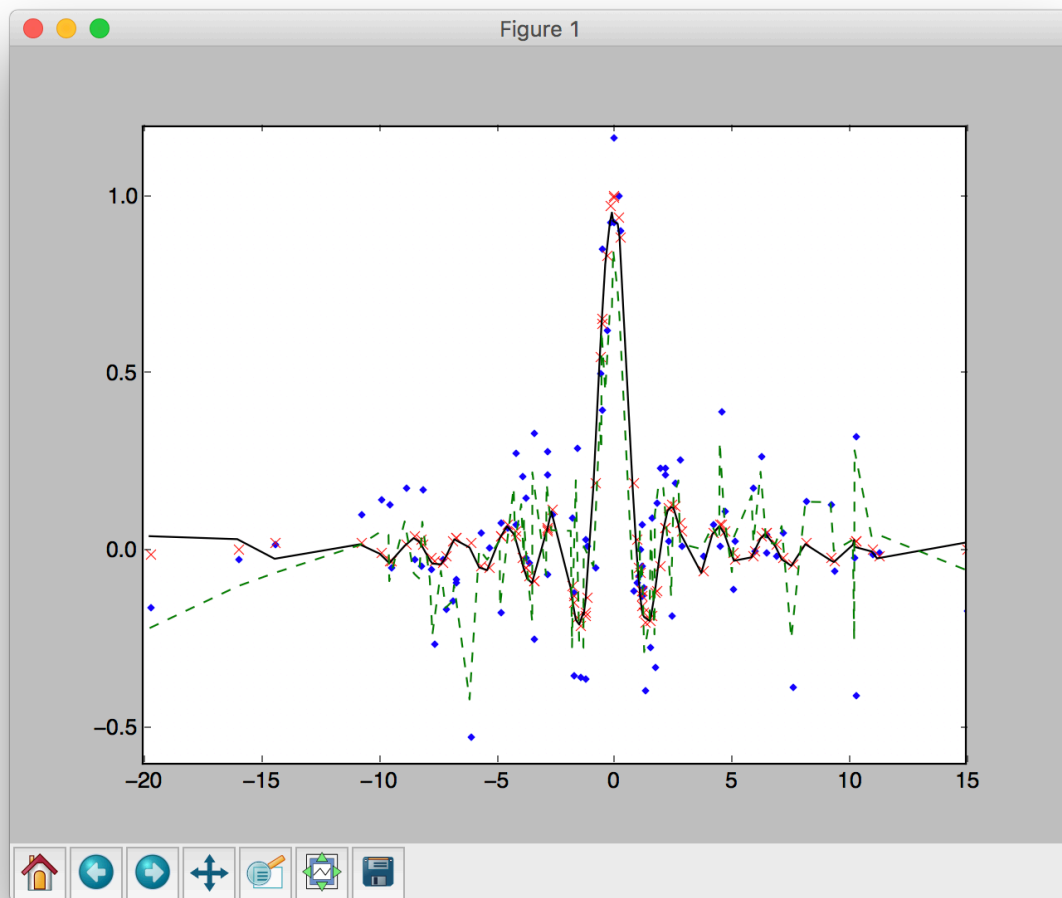


Problem 1:

The program which implements 10-fold cross-validation approach to find the best set of hyper-parameters C , ε (epsilon) and γ (gamma) in a Support Vector Machine for Regression (SVR) is downloaded.

The program which generates data points of the “sinc” function contaminated with random noise is downloaded and given the file name finalGenData.py

The program for samples $N=100$ is run and the output plot along with the best hyper-parameters are obtained.



The downloaded program is run and the hyper-parameters are obtained for every 'Score' which is better than the 'bestScore' i.e., for $\text{Score} > \text{bestScore}$, the best values of C , ϵ (epsilon) and γ (gamma) are obtained.

```
Final — -bash — 94x28
The following packages will be UPDATED:
    conda: 4.2.9-py27_0 --> 4.2.13-py27_0
    scikit-learn: 0.17.1-np111py27_2 --> 0.18.1-np111py27_0

Proceed ([y]/n)? y

Fetching packages ...
conda-env-2.6. 100% |#####| Time: 0:00:00 581.76 kB/s
conda-4.2.13-p 100% |#####| Time: 0:00:00 2.60 MB/s
scikit-learn-0 100% |#####| Time: 0:00:00 12.45 MB/s
Extracting packages ...
[ COMPLETE ] |#####| 100%
Unlinking packages ...
[ COMPLETE ] |#####| 100%
Linking packages ...
[ COMPLETE ] |#####| 100%
Anirudhs-MacBook-Pro:Final Anirudh$ python final.SVM.sinc.py
C 0.03125, epsilon 0.0, gamma 3.0517578125e-05. Testing set CV score: -1.548211
C 0.03125, epsilon 0.0, gamma 8.0. Testing set CV score: -1.346822
C 8.0, epsilon 0.0, gamma 0.5. Testing set CV score: -0.906322
[LibSVM] .....
optimization finished, #iter = 7269
obj = -3.535969, rho = -0.042700
nSV = 100, nBSV = 13
Training set score: 0.999801
Testing set score: 0.997579
Anirudhs-MacBook-Pro:Final Anirudh$
```

After all the iterations are done, the Training set score, Testing set score are calculated and obtained which have a negligible difference which shows the accuracy of the system.

Now, the program is modified and conducted for the samples $N=1000$.

The best hyper-parameters are noted for a highest Testing CV score as follows

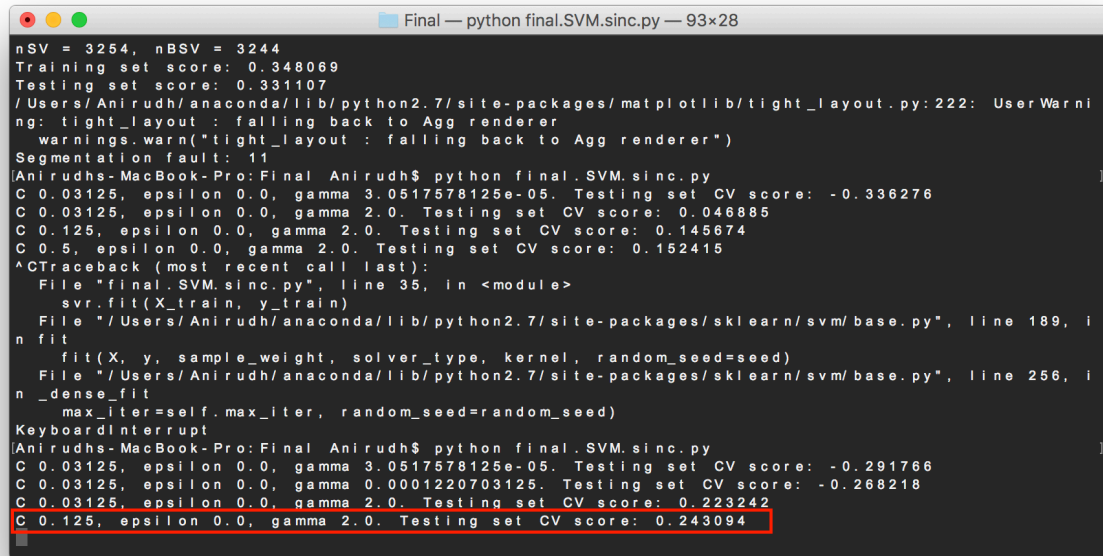
```
Final — -bash — 94x28
Anirudhs-MacBook-Pro:Final Anirudh$ python final.SVM.sinc.py
C 0.03125, epsilon 0.0, gamma 3.0517578125e-05. Testing set CV score: -0.255180
C 0.03125, epsilon 0.0, gamma 0.03125. Testing set CV score: -0.237875
C 0.03125, epsilon 0.0, gamma 2.0. Testing set CV score: -0.086164
C 0.125, epsilon 0.0, gamma 2.0. Testing set CV score: -0.037331
C 0.5, epsilon 0.0, gamma 2.0. Testing set CV score: -0.002925
[LibSVM] .....
Warning: using -h 0 may be faster
*
optimization finished, #iter = 7520
obj = -1.284918, rho = -0.035584
nSV = 807, nBSV = 145
Training set score: 0.999940
Testing set score: 0.999830
Segmentation fault: 11
Anirudhs-MacBook-Pro:Final Anirudh$
Anirudhs-MacBook-Pro:Final Anirudh$
Anirudhs-MacBook-Pro:Final Anirudh$
```

These values are posted in the google forms as provided.

Now, the same program is modified and conducted and conducted for samples $N=10000$.

The program was run for almost five hours but it didn't give out the answer. So, the latest value of bestScore hyper-parameters are considered as the best parameters.

The results obtained were



```
nSV = 3254, nBSV = 3244
Training set score: 0.348069
Testing set score: 0.331107
/Users/Anirudh/anaconda/lib/python2.7/site-packages/matplotlib/tight_layout.py:222: UserWarning: tight_layout : falling back to Agg renderer
  warnings.warn("tight_layout : falling back to Agg renderer")
Segmentation fault: 11
[Anirudhs-MacBook-Pro:Final Anirudh$ python final.SVM.sinc.py
C 0.03125, epsilon 0.0, gamma 3.0517578125e-05. Testing set CV score: -0.336276
C 0.03125, epsilon 0.0, gamma 2.0. Testing set CV score: 0.046885
C 0.125, epsilon 0.0, gamma 2.0. Testing set CV score: 0.145674
C 0.5, epsilon 0.0, gamma 2.0. Testing set CV score: 0.152415
^CTraceback (most recent call last):
  File "final.SVM.sinc.py", line 35, in <module>
    svr.fit(X_train, y_train)
  File "/Users/Anirudh/anaconda/lib/python2.7/site-packages/sklearn/svm/base.py", line 189, in fit
    fit(X, y, sample_weight, solver_type, kernel, random_seed=seed)
  File "/Users/Anirudh/anaconda/lib/python2.7/site-packages/sklearn/svm/base.py", line 256, in _dense_fit
    max_iter=self.max_iter, random_seed=random_seed)
KeyboardInterrupt
[Anirudhs-MacBook-Pro:Final Anirudh$ python final.SVM.sinc.py
C 0.03125, epsilon 0.0, gamma 3.0517578125e-05. Testing set CV score: -0.291766
C 0.03125, epsilon 0.0, gamma 0.0001220703125. Testing set CV score: -0.268218
C 0.03125, epsilon 0.0, gamma 2.0. Testing set CV score: 0.223242
C 0.125, epsilon 0.0, gamma 2.0. Testing set CV score: 0.243094
```

As we can observe, as we are increasing the number of samples, the best values of C is considerably reducing to a very smaller number, ϵ (epsilon) remained constant and γ (gamma) increased with increasing Training set CV score. Also, the as we increase the number of samples, the Training set score and Testing set scores are converging to a greater extent.

Problem 2:

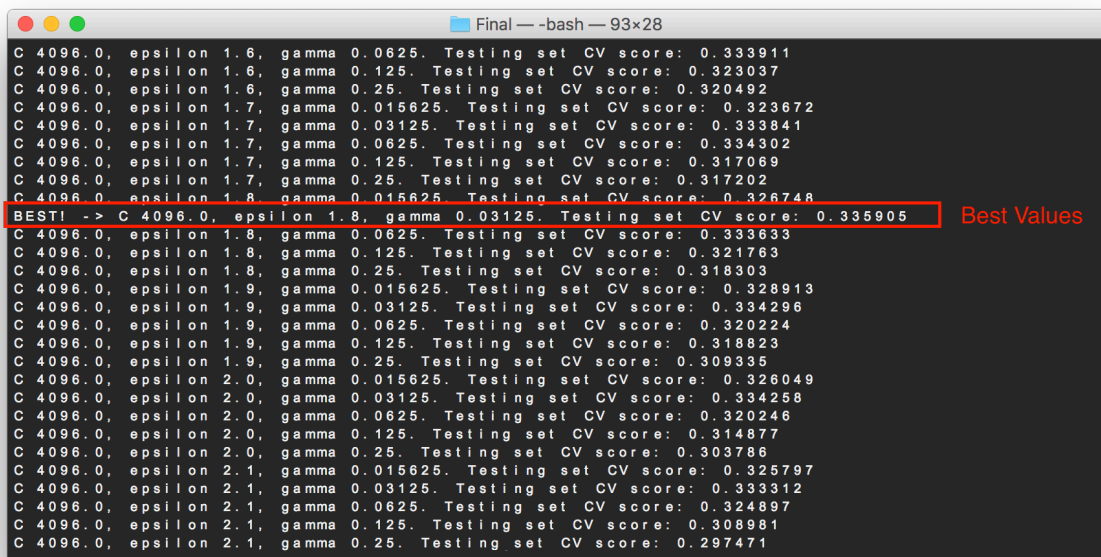
The given program was coded and saved with a file name finalGetDigits.py

The program which implements 10-fold cross-validation approach to find the best set of hyper-parameters C , ϵ (epsilon) and γ (gamma) in a Support Vector Machine for Regression (SVR) is downloaded.

The downloaded program is run and iterations are seen for increasing values of hyper parameters C (power of 2s from 2^6), ϵ (adding 0.1 from 0.5 to 2.5) and γ (twice from 0.015625 to 0.25) and the values of Testing set CV score are noted.

Also, the best values of the hyper parameters (best C , best ϵ and best γ) are obtained for the best value of Testing set CV score (where a maxima occurs)

For example,



```
Final -- -bash -- 93x28
C 4096.0, epsilon 1.6, gamma 0.0625. Testing set CV score: 0.333911
C 4096.0, epsilon 1.6, gamma 0.125. Testing set CV score: 0.323037
C 4096.0, epsilon 1.6, gamma 0.25. Testing set CV score: 0.320492
C 4096.0, epsilon 1.7, gamma 0.015625. Testing set CV score: 0.323672
C 4096.0, epsilon 1.7, gamma 0.03125. Testing set CV score: 0.333841
C 4096.0, epsilon 1.7, gamma 0.0625. Testing set CV score: 0.334302
C 4096.0, epsilon 1.7, gamma 0.125. Testing set CV score: 0.317069
C 4096.0, epsilon 1.7, gamma 0.25. Testing set CV score: 0.317202
C 4096.0, epsilon 1.8, gamma 0.015625. Testing set CV score: 0.326748
BEST! -> C 4096.0, epsilon 1.8, gamma 0.03125. Testing set CV score: 0.335905
C 4096.0, epsilon 1.8, gamma 0.0625. Testing set CV score: 0.333633
C 4096.0, epsilon 1.8, gamma 0.125. Testing set CV score: 0.321763
C 4096.0, epsilon 1.8, gamma 0.25. Testing set CV score: 0.318303
C 4096.0, epsilon 1.9, gamma 0.015625. Testing set CV score: 0.328913
C 4096.0, epsilon 1.9, gamma 0.03125. Testing set CV score: 0.334296
C 4096.0, epsilon 1.9, gamma 0.0625. Testing set CV score: 0.320224
C 4096.0, epsilon 1.9, gamma 0.125. Testing set CV score: 0.318823
C 4096.0, epsilon 1.9, gamma 0.25. Testing set CV score: 0.309335
C 4096.0, epsilon 2.0, gamma 0.015625. Testing set CV score: 0.326049
C 4096.0, epsilon 2.0, gamma 0.03125. Testing set CV score: 0.334258
C 4096.0, epsilon 2.0, gamma 0.0625. Testing set CV score: 0.320246
C 4096.0, epsilon 2.0, gamma 0.125. Testing set CV score: 0.314877
C 4096.0, epsilon 2.0, gamma 0.25. Testing set CV score: 0.303786
C 4096.0, epsilon 2.1, gamma 0.015625. Testing set CV score: 0.325797
C 4096.0, epsilon 2.1, gamma 0.03125. Testing set CV score: 0.333312
C 4096.0, epsilon 2.1, gamma 0.0625. Testing set CV score: 0.324897
C 4096.0, epsilon 2.1, gamma 0.125. Testing set CV score: 0.308981
C 4096.0, epsilon 2.1, gamma 0.25. Testing set CV score: 0.297471
```

The best values (hyper-parameters) are obtained for every 'Score' which is better than the 'bestScore' i.e., for Score > bestScore, the best values of C , ϵ (epsilon) and γ (gamma) are obtained.

When ever a maxima is seen in Training set CV score, best values are noted.

The final output of the program after all the 51267 iterations is

```
C 4096.0, epsilon 2.3, gamma 0.125. Testing set CV score: 0.305671  
C 4096.0, epsilon 2.3, gamma 0.25. Testing set CV score: 0.294358  
C 4096.0, epsilon 2.4, gamma 0.015625. Testing set CV score: 0.321655  
C 4096.0, epsilon 2.4, gamma 0.03125. Testing set CV score: 0.306296  
C 4096.0, epsilon 2.4, gamma 0.0625. Testing set CV score: 0.317836  
C 4096.0, epsilon 2.4, gamma 0.125. Testing set CV score: 0.298992  
C 4096.0, epsilon 2.4, gamma 0.25. Testing set CV score: 0.290367  
C 4096.0, epsilon 2.5, gamma 0.015625. Testing set CV score: 0.324188  
C 4096.0, epsilon 2.5, gamma 0.03125. Testing set CV score: 0.316030  
C 4096.0, epsilon 2.5, gamma 0.0625. Testing set CV score: 0.308841  
C 4096.0, epsilon 2.5, gamma 0.125. Testing set CV score: 0.290031  
C 4096.0, epsilon 2.5, gamma 0.25. Testing set CV score: 0.274304
```

[LibSVM].....*

.....*

.....*

.....*

.....*

.....*

optimization finished, #iter = 511267
obj = -19551733.014279, rho = -87.625466
nSV = 3254, nBSV = 3244
Training set score: 0.348069
Testing set score: 0.331107

```
/Users/Anirudh/anaconda/lib/python2.7/site-packages/matplotlib/tight_layout.py:222: UserWarning:  
ng: tight_layout : falling back to Agg renderer  
    warnings.warn("tight_layout : falling back to Agg renderer")  
Segmentation fault: 11  
Anirudhs-MacBook-Pro:Final Anirudh$
```