Jae Park
Data Mining HW 5
Dual SVM

A good C value and a kernel parameter (spread) were found using sklearn.svm.SVC:

$C = 10, \sigma^2 = 1$

maxiter (max iteration permitted): 1000 was chosen for speed

----------< **Linear** >----------

C = 0.01

```
jaebuntu@jaebuntu ~/Dropbox/rpi/Fall2021/Data-Mining/hw5 master ± python Assign5.py energydata_complete.csv 0.01 0.001 1000 linear 1
SVM converged after 105 iterations
Validation set accuracy: 67.44%
Test set accuracy: 69.91%
```

C = 0.1

```
jaebuntu@jaebuntu ~/Dropbox/rpi/Fall2021/Data-Mining/hw5 master ± python Assign5.py energydata_complete.csv 0.1 0.001 1000 linear 1
SVM converged after 624 iterations
Validation set accuracy: 69.29%
Test set accuracy: 68.21%
```

C = 1

```
jaebuntu@jaebuntu ~/Dropbox/rpi/Fall2021/Data-Mining/hw5 master ± python Assign5.py energydata_complete.csv 1 0.001 1000 linear 1
SVM converged after 1000 iterations
Validation set accuracy: 69.84%
Test set accuracy: 67.22%
```

C = 10 <Best Validation set accuracy>

```
jaebuntu@jaebuntu ~/Dropbox/rpi/Fall2021/Data-Mining/hw5 master ± python Assign5.py energydata_complete.csv 10 0.001 1000 linear 1
SVM converged after 1000 iterations
Validation set accuracy: 69.86%
Test set accuracy: 67.01%
```

C = 11

```
jaebuntu@jaebuntu ~/Dropbox/rpi/Fall2021/Data-Mining/hw5 master ± python Assign5.py energydata_complete.csv 11 0.001 1000 linear 1
SVM converged after 1000 iterations
(2000, 5000)
Validation set accuracy: 69.26%
Test set accuracy: 67.35%
```

----------< **Gaussian** >----------

$\sigma^2 = 0.1$

```
jaebuntu@jaebuntu ~/Dropbox/rpi/Fall2021/Data-Mining/hw5 master ± python Assign5.py energydata_complete.csv 10 0.001 1000 gaussian 0.1
SVM converged after 301 iterations
Validation set accuracy: 59.18%
Test set accuracy: 61.97%
```

$\sigma^2 = 1$

```
jaebuntu@jaebuntu ~/Dropbox/rpi/Fall2021/Data-Mining/hw5 master ± python Assign5.py energydata_complete.csv 10 0.001 1000 gaussian 1
SVM converged after 1000 iterations
Validation set accuracy: 62.52%
Test set accuracy: 62.51%
```

$\sigma^2 = 2$

```
jaebuntu@jaebuntu ~/Dropbox/rpi/Fall2021/Data-Mining/hw5 master ± python Assign5.py energydata_complete.csv 10 0.001 1000 gaussian 2
SVM converged after 1000 iterations
Validation set accuracy: 62.18%
Test set accuracy: 64.96%
```

$\sigma^2 = 3$

```
jaebuntu@jaebuntu ~/Dropbox/rpi/Fall2021/Data-Mining/hw5 master ± python Assign5.py energydata_complete.csv 10 0.001 1000 gaussian 3
SVM converged after 1000 iterations
Validation set accuracy: 60.83%
Test set accuracy: 65.65%
```

$\sigma^2 = 10$

```
jaebuntu@jaebuntu ~/Dropbox/rpi/Fall2021/Data-Mining/hw5 master ± python Assign5.py energydata_complete.csv 10 0.001 1000 gaussian 10
SVM converged after 1000 iterations
Validation set accuracy: 61.38%
Test set accuracy: 67.38%
```

Final output of the **weight vector** and **bias** with **C** = 10, σ^2 = 1

$ python3 Assign5.py energydata_complete.csv 10 0.001 1000 **linear** 1

```
weight vector:
[-0.039 -0.89   1.255 -2.542 -1.599  2.832 -0.217  0.031  0.099 -0.16
  0.133 -0.149 -0.851 -0.456  0.336 -0.433 -0.724  1.403  0.251  0.319
  0.484  0.209 -0.108  0.064 -0.087 -0.254  0.004]
bias:
-0.03924990468010492
Number of Support Vectors: 749
```

weight vector = (27 x 1)