ESE 546, FALL 2020

HOMEWORK 0

ARJUN GOVIND [AGOVIND@WHARTON],
COLLABORATORS: SHEIL SARDA [SHEILS@SEAS]

Solution 1 (Time spent: 1 hour)**. 1(a)**: Since we are adding a slack term, our new goal will be to ensure that the total ξ_i values are minimized in addition to our original constraint. As such, our new optimization problem becomes:

minimize $\frac{1}{2}||\theta||^2 + \sum_i \xi_i$ subject to $y_i(\theta^T x_i + \theta_0) >= 1$

1(b): In SVM, support samples refer to the number of samples in each class for the training data.

1(c): Please see enclosed notebook for the code. The reason you would initially want to construct just the training and validation dataset is because our test dataset should be completely separate from the training and validation data–our model should not be built or tuned based on that data. Instead, we use the training dataset to train our model and the validation dataset to tune hyperparameters.

1(d): Please see enclosed notebook for the code and output; it has been copied below. The C parameter is a penalty (regularization) parameter. A higher value of C increases in-sample accuracy but also increases the risk of overfitting. The gamma parameter is the kernel parameter, denoting which kernel function you want to use. The validation accuracy was 0.96, and the confusion matrix is below.

```
Validation Accuracy: 0.9620714285714286
array([[1327,    1,    2,    0,    2,    2,    3,    1,    4,    1],
       [   0, 1573,    7,    6,    1,    1,    0,    5,    6,    1],
       [   1,    6, 1329,    6,    9,    4,    5,    8,    9,    3],
       [   1,    5,   15, 1351,    1,   22,    4,   14,   13,    7],
       [   3,    1,    3,    0, 1243,    0,    5,    3,    2,   35],
       [   2,    6,    3,   28,    3, 1217,    9,    0,    5,    0],
       [   5,    2,    0,    0,    8,    8, 1371,    0,    2,    0],
       [   2,    8,   12,    1,    7,    0,    0, 1447,    4,   22],
       [   2,    8,    8,   21,   11,   12,    7,    7, 1279,    2],
       [   6,    7,    3,   19,   27,    2,    0,   19,    5, 1332]])
```

1(e): Based on the svm.SVC documentation and user guide, the main parameters I was unfamiliar with were ”shrinking”,”decision-function-shape”, and ”tol”. The ”shrinking” parameter is a boolean that indicates whether to use the Shrinking Heuristic, which reduces training time especially when the number of iterations is large. Based on the user guide, the optimization that the SVC is performing is:

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i$$
$$\text{subject to } y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i,$$
$$\zeta_i \geq 0, i = 1, \ldots, n$$

1(f): The sklearn implementation deals with multiclass classification with a "one-versus-one" approach. In total, $n_{classes} * (n_{classes} - 1)/2$ classifiers are constructed and each one trains data from two classes. It also provides the option to transform the results of the "one-versus-one" classifiers to a "one-vs-rest" decision function of shape (n-samples, n-classes).

1(g): Please see enclosed notebook for the code and output; this has been copied below.

```
Validation accuracy of 0.57370 with {'C': 0.01}
Validation accuracy of 0.91420 with {'C': 0.1}
Validation accuracy of 0.95240 with {'C': 1}
Validation accuracy of 0.95820 with {'C': 10}
Validation accuracy of 0.95720 with {'C': 100}
Best C value:
{'C': 10}
```

1(h) and 1(j): Please see enclosed notebook for the code and output.

Solution 2 (Time spent: 1 hour)**.** Your solution goes here.

Solution 3 (Time spent: 1 hour)**.**