HW5_Myunghee_ID_2446752777

March 27, 2019

1 HW5, Myunghee Lee (USC ID: 2446752777)

1. Multi-class and Multi-Label Classication Using Support Vector Machines

(a) Choose 70% of the data randomly as the training set.

```
In [1]: import pandas as pd
        from sklearn.model_selection import train_test_split  # to divide training and test da

        df = pd.read_csv("Frogs_MFCCs.csv")
        print(df)
```

```
      MFCCs_ 1  MFCCs_ 2  MFCCs_ 3  MFCCs_ 4  MFCCs_ 5  MFCCs_ 6  MFCCs_ 7  \
0          1.0  0.152936 -0.105586  0.200722  0.317201  0.260764  0.100945
1          1.0  0.171534 -0.098975  0.268425  0.338672  0.268353  0.060835
2          1.0  0.152317 -0.082973  0.287128  0.276014  0.189867  0.008714
3          1.0  0.224392  0.118985  0.329432  0.372088  0.361005  0.015501
4          1.0  0.087817 -0.068345  0.306967  0.330923  0.249144  0.006884
5          1.0  0.099704 -0.033408  0.349895  0.344535  0.247569  0.022407
6          1.0  0.021676 -0.062075  0.318229  0.380439  0.179043 -0.041667
7          1.0  0.145130 -0.033660  0.284166  0.279537  0.175211  0.005791
8          1.0  0.271326  0.027777  0.375738  0.385432  0.272457  0.098192
9          1.0  0.120565 -0.107235  0.316555  0.364437  0.307757  0.025992
10         1.0  0.148539 -0.096910  0.257523  0.260881  0.312603  0.134134
11         1.0  0.277948  0.091657  0.331656  0.307372  0.257359  0.065702
12         1.0  0.106109 -0.025790  0.358875  0.297543  0.244335  0.016446
13         1.0  0.126523 -0.040482  0.341129  0.381446  0.261154 -0.017049
14         1.0  0.267687  0.099327  0.510454  0.511468  0.317788  0.067992
15         1.0  0.137623 -0.085808  0.322446  0.344695  0.285642  0.056517
16         1.0  0.263944  0.090358  0.368888  0.356645  0.252806  0.063921
17         1.0  0.146299 -0.075174  0.291935  0.367094  0.268947  0.054049
18         1.0  0.179298 -0.038306  0.319636  0.383029  0.275313  0.099083
19         1.0  0.273218 -0.234703 -0.079620  0.159811  0.416406  0.368838
20         1.0  0.196429  0.009021  0.317772  0.293484  0.185684  0.044063
21         1.0  0.230999  0.135657  0.431966  0.403423  0.276571  0.060464
22         1.0  0.145109 -0.035846  0.282707  0.291044  0.206862  0.048627
23         1.0  0.235682  0.029241  0.349117  0.355932  0.290697  0.081008
```

```
24          1.0  0.146944 -0.009583  0.352534  0.313435  0.197599  0.015352
25          1.0  0.233512  0.067249  0.352310  0.316899  0.220584  0.044433
26          1.0  0.172672 -0.037870  0.301100  0.303533  0.203767  0.017798
27          1.0  0.198494  0.078718  0.478231  0.425219  0.257916  0.049410
28          1.0  0.165998 -0.004175  0.289963  0.295084  0.224001  0.085586
29          1.0  0.155225 -0.063337  0.231699  0.284514  0.219596  0.038581
...         ...       ...       ...       ...       ...       ...       ...
7165        1.0  0.132365  0.503936  0.271392 -0.042392  0.024487  0.080317
7166        1.0  0.165383  0.408082  0.270187  0.015300  0.019000  0.025284
7167        1.0  0.411916  0.322796  0.344183  0.333873  0.094867 -0.230982
7168        1.0  0.404936  0.726247  0.167376 -0.169260  0.112154  0.050805
7169        1.0  0.209224  0.625373  0.246565 -0.108078  0.174846  0.100347
7170        1.0  0.242842  0.516606  0.263976 -0.040676  0.052257  0.055201
7171        1.0  0.153789  0.522557  0.267617 -0.043499  0.034436  0.081305
7172        1.0  0.266711  0.735734  0.233772 -0.075867  0.140193  0.080212
7173        1.0  0.183156  0.506970  0.247162 -0.048503  0.041090  0.093071
7174        1.0  0.197679  0.553775  0.269271 -0.056261  0.062528  0.061367
7175        1.0 -0.673025 -0.279960  0.065945  0.086890  0.469024  0.088019
7176        1.0 -0.574239 -0.258302  0.007644  0.078566  0.402132  0.093501
7177        1.0 -0.515567 -0.219812  0.118473  0.106919  0.471155  0.121196
7178        1.0 -0.614409 -0.261315  0.051138  0.062119  0.501500  0.096506
7179        1.0 -0.578837 -0.331501  0.105921  0.089931  0.456617  0.079131
7180        1.0 -0.528595 -0.208051  0.103669  0.086537  0.408476  0.069610
7181        1.0 -0.442139 -0.328404  0.031452  0.056017  0.424856  0.073288
7182        1.0 -0.616029 -0.302357  0.063417  0.095671  0.439930  0.069414
7183        1.0 -0.547168 -0.266780  0.056115  0.048947  0.423631  0.081924
7184        1.0 -0.520958 -0.258779 -0.070416 -0.025129  0.447967  0.180033
7185        1.0 -0.512794  0.056322  0.259677  0.030140  0.369783 -0.117154
7186        1.0 -0.591520 -0.268901  0.050042  0.116960  0.444706  0.059268
7187        1.0 -0.507564 -0.249969  0.031781 -0.079888  0.484274  0.125143
7188        1.0 -0.512599 -0.171956  0.325813  0.169600  0.421567 -0.123749
7189        1.0 -0.558546 -0.238442  0.066527  0.123090  0.395953  0.066522
7190        1.0 -0.554504 -0.337717  0.035533  0.034511  0.443451  0.093889
7191        1.0 -0.517273 -0.370574  0.030673  0.068097  0.402890  0.096628
7192        1.0 -0.582557 -0.343237  0.029468  0.064179  0.385596  0.114905
7193        1.0 -0.519497 -0.307553 -0.004922  0.072865  0.377131  0.086866
7194        1.0 -0.508833 -0.324106  0.062068  0.078211  0.397188  0.094596

      MFCCs_ 8  MFCCs_ 9  MFCCs_10  ...  MFCCs_17  MFCCs_18  MFCCs_19  \
0    -0.150063 -0.171128  0.124676  ... -0.108351 -0.077623 -0.009568
1    -0.222475 -0.207693  0.170883  ... -0.090974 -0.056510 -0.035303
2    -0.242234 -0.219153  0.232538  ... -0.050691 -0.023590 -0.066722
3    -0.194347 -0.098181  0.270375  ... -0.136009 -0.177037 -0.130498
4    -0.265423 -0.172700  0.266434  ... -0.048885 -0.053074 -0.088550
5    -0.213767 -0.127916  0.277353  ... -0.080487 -0.130089 -0.171478
6    -0.252300 -0.167117  0.220027  ... -0.046620 -0.055146 -0.085972
7    -0.183329 -0.158483  0.192567  ... -0.055978 -0.048219 -0.056637
8    -0.173730 -0.157857  0.207181  ... -0.120723 -0.112607 -0.156933
```

9	−0.294179	−0.223236	0.268435	...	−0.051073	−0.052568	−0.111338
10	−0.216262	−0.189334	0.261960	...	−0.034082	−0.120716	−0.100800
11	−0.191860	−0.133537	0.220020	...	−0.119167	−0.110900	−0.112485
12	−0.288733	−0.146731	0.314207	...	−0.062939	−0.071182	−0.066827
13	−0.294064	−0.222278	0.282338	...	−0.071544	−0.060630	−0.067230
14	−0.202826	−0.142236	0.235510	...	−0.138830	−0.139922	−0.126448
15	−0.314418	−0.252324	0.288897	...	−0.058694	−0.072913	−0.064263
16	−0.155007	−0.137743	0.200262	...	−0.074168	−0.083995	−0.104413
17	−0.242952	−0.232617	0.235722	...	−0.051154	−0.038580	−0.022396
18	−0.207998	−0.219215	0.182845	...	−0.110969	−0.105833	−0.115237
19	0.016878	−0.171288	−0.115424	...	−0.253103	−0.154244	−0.002606
20	−0.169936	−0.121461	0.237437	...	−0.045409	−0.067118	−0.047625
21	−0.192200	−0.187348	0.180486	...	−0.163367	−0.170739	−0.169508
22	−0.172111	−0.115698	0.251256	...	−0.044077	−0.067219	−0.058514
23	−0.193793	−0.151462	0.212130	...	−0.158932	−0.098565	−0.078413
24	−0.216492	−0.133865	0.278309	...	−0.043595	−0.052536	−0.024106
25	−0.172653	−0.127641	0.190017	...	−0.105466	−0.070941	−0.085853
26	−0.197260	−0.124021	0.256757	...	−0.078277	−0.023172	0.002809
27	−0.146034	−0.127461	0.170725	...	−0.172904	−0.164822	−0.152341
28	−0.180608	−0.169064	0.221722	...	−0.067536	−0.084953	−0.116397
29	−0.183926	−0.108442	0.245208	...	−0.072937	−0.018316	−0.034315
...	...	...	...	...	...	...	...
7165	0.112706	−0.071058	−0.114414	...	−0.006677	−0.070338	0.003964
7166	0.120438	0.042605	−0.053960	...	0.007785	−0.002231	0.020593
7167	0.276750	0.232506	−0.372219	...	0.029924	0.031286	−0.038367
7168	0.078670	−0.243258	−0.120933	...	0.090483	0.126233	0.015162
7169	0.035056	−0.149125	−0.099348	...	−0.041610	−0.039468	0.086045
7170	0.125361	−0.023455	−0.064192	...	0.004518	−0.013307	−0.001609
7171	0.113272	−0.034906	−0.055563	...	−0.005648	−0.028727	0.031281
7172	0.119675	−0.038130	−0.033289	...	−0.027436	−0.007999	0.058854
7173	0.109081	−0.049473	−0.044991	...	−0.008959	−0.034239	0.036489
7174	0.094837	−0.047792	−0.063015	...	0.021719	−0.041396	−0.013619
7175	−0.197498	0.031180	0.096919	...	0.052054	0.114138	0.069353
7176	−0.146350	0.040779	0.105841	...	0.062850	0.038051	0.019184
7177	−0.136494	0.027486	0.106110	...	0.049024	0.109761	0.043604
7178	−0.155093	0.035657	0.121933	...	0.048454	0.072420	0.017823
7179	−0.133406	0.056485	0.093265	...	0.079075	0.125759	0.044132
7180	−0.155217	0.080287	0.127136	...	0.053063	0.085869	0.020705
7181	−0.140148	0.043070	0.087675	...	0.024943	0.068279	0.015953
7182	−0.145534	0.031354	0.069073	...	0.056971	0.037285	0.023818
7183	−0.184252	0.024682	0.111803	...	0.049831	0.115522	0.051333
7184	−0.062297	0.032410	−0.005379	...	0.027392	0.019266	−0.057772
7185	−0.189292	0.248948	0.193566	...	0.096589	0.255261	0.047039
7186	−0.158389	0.066648	0.083924	...	0.074444	0.117996	0.045522
7187	−0.069555	0.114265	0.099647	...	0.011524	0.111998	0.015886
7188	−0.298284	0.089382	0.243902	...	0.021225	0.157321	0.042847
7189	−0.152216	0.078294	0.094184	...	0.034202	0.040540	−0.003755
7190	−0.100753	0.037087	0.081075	...	0.069430	0.071001	0.021591

```
7191 -0.116460  0.063727  0.089034  ...  0.061127  0.068978  0.017745
7192 -0.103317  0.070370  0.081317  ...  0.082474  0.077771 -0.009688
7193 -0.115799  0.056979  0.089316  ...  0.051796  0.069073  0.017963
7194 -0.117672  0.058874  0.076180  ...  0.061455  0.072983 -0.003980

      MFCCs_20  MFCCs_21  MFCCs_22          Family      Genus  \
0     0.057684  0.118680  0.014038  Leptodactylidae  Adenomera
1     0.020140  0.082263  0.029056  Leptodactylidae  Adenomera
2    -0.025083  0.099108  0.077162  Leptodactylidae  Adenomera
3    -0.054766 -0.018691  0.023954  Leptodactylidae  Adenomera
4    -0.031346  0.108610  0.079244  Leptodactylidae  Adenomera
5    -0.071569  0.077643  0.064903  Leptodactylidae  Adenomera
6    -0.009127  0.065630  0.044040  Leptodactylidae  Adenomera
7    -0.022419  0.070085  0.021419  Leptodactylidae  Adenomera
8    -0.118527 -0.002471  0.002304  Leptodactylidae  Adenomera
9    -0.040014  0.090204  0.088025  Leptodactylidae  Adenomera
10   -0.001992  0.111462  0.103637  Leptodactylidae  Adenomera
11   -0.053184  0.044291 -0.011456  Leptodactylidae  Adenomera
12   -0.028048  0.058353  0.064368  Leptodactylidae  Adenomera
13   -0.038196  0.070127  0.048440  Leptodactylidae  Adenomera
14   -0.067570  0.057888 -0.011998  Leptodactylidae  Adenomera
15    0.022455  0.130752  0.074132  Leptodactylidae  Adenomera
16   -0.071431  0.028842  0.019180  Leptodactylidae  Adenomera
17   -0.018891  0.051480  0.031871  Leptodactylidae  Adenomera
18   -0.012819  0.083194  0.052101  Leptodactylidae  Adenomera
19    0.092999  0.091724  0.003595  Leptodactylidae  Adenomera
20   -0.005875  0.053107  0.030669  Leptodactylidae  Adenomera
21   -0.112446  0.065072  0.050254  Leptodactylidae  Adenomera
22   -0.001358  0.082058  0.045492  Leptodactylidae  Adenomera
23   -0.043410  0.033478 -0.006953  Leptodactylidae  Adenomera
24    0.016081  0.062506  0.042285  Leptodactylidae  Adenomera
25   -0.025053  0.055194  0.003098  Leptodactylidae  Adenomera
26    0.009133  0.063916  0.047634  Leptodactylidae  Adenomera
27   -0.068851  0.024412 -0.009821  Leptodactylidae  Adenomera
28   -0.008499  0.101151  0.022322  Leptodactylidae  Adenomera
29   -0.029563  0.051715  0.023925  Leptodactylidae  Adenomera
...        ...       ...       ...              ...        ...
7165  0.032534  0.002975  0.002724          Hylidae     Scinax
7166  0.014499  0.008208  0.018023          Hylidae     Scinax
7167 -0.007837  0.102445  0.000755          Hylidae     Scinax
7168  0.059531  0.050905 -0.026956          Hylidae     Scinax
7169  0.071944 -0.021574  0.036031          Hylidae     Scinax
7170  0.022170  0.014327  0.003001          Hylidae     Scinax
7171  0.033687  0.005242  0.014982          Hylidae     Scinax
7172  0.008365  0.019086  0.090131          Hylidae     Scinax
7173  0.042205  0.003017  0.005602          Hylidae     Scinax
7174  0.010337  0.017903  0.022108          Hylidae     Scinax
7175  0.050519 -0.069281 -0.106642          Hylidae     Scinax
```

```
7176  0.027745 -0.109677 -0.131287          Hylidae     Scinax
7177  0.047369 -0.018168 -0.085380          Hylidae     Scinax
7178 -0.021418 -0.094280 -0.091600          Hylidae     Scinax
7179  0.017652 -0.077831 -0.132563          Hylidae     Scinax
7180  0.009959 -0.072003 -0.121693          Hylidae     Scinax
7181  0.058530 -0.009746 -0.080940          Hylidae     Scinax
7182  0.042104 -0.035603 -0.097343          Hylidae     Scinax
7183  0.019697 -0.061764 -0.078648          Hylidae     Scinax
7184  0.037710  0.047717  0.001436          Hylidae     Scinax
7185 -0.016643 -0.046598 -0.055862          Hylidae     Scinax
7186  0.050734 -0.034757 -0.085131          Hylidae     Scinax
7187  0.008636 -0.021933 -0.069692          Hylidae     Scinax
7188  0.006852  0.005439 -0.013693          Hylidae     Scinax
7189  0.036059 -0.031853 -0.090206          Hylidae     Scinax
7190  0.052449 -0.021860 -0.079860          Hylidae     Scinax
7191  0.046461 -0.015418 -0.101892          Hylidae     Scinax
7192  0.027834 -0.000531 -0.080425          Hylidae     Scinax
7193  0.041803 -0.027911 -0.096895          Hylidae     Scinax
7194  0.031560 -0.029355 -0.087910          Hylidae     Scinax
```

```
             Species  RecordID
0     AdenomeraAndre         1
1     AdenomeraAndre         1
2     AdenomeraAndre         1
3     AdenomeraAndre         1
4     AdenomeraAndre         1
5     AdenomeraAndre         1
6     AdenomeraAndre         1
7     AdenomeraAndre         1
8     AdenomeraAndre         1
9     AdenomeraAndre         1
10    AdenomeraAndre         1
11    AdenomeraAndre         1
12    AdenomeraAndre         1
13    AdenomeraAndre         1
14    AdenomeraAndre         1
15    AdenomeraAndre         1
16    AdenomeraAndre         1
17    AdenomeraAndre         1
18    AdenomeraAndre         1
19    AdenomeraAndre         1
20    AdenomeraAndre         1
21    AdenomeraAndre         1
22    AdenomeraAndre         1
23    AdenomeraAndre         1
24    AdenomeraAndre         1
25    AdenomeraAndre         1
26    AdenomeraAndre         1
```

```
27     AdenomeraAndre        1
28     AdenomeraAndre        1
29     AdenomeraAndre        1
...               ...      ...
7165      ScinaxRuber       59
7166      ScinaxRuber       59
7167      ScinaxRuber       59
7168      ScinaxRuber       59
7169      ScinaxRuber       59
7170      ScinaxRuber       59
7171      ScinaxRuber       59
7172      ScinaxRuber       59
7173      ScinaxRuber       59
7174      ScinaxRuber       59
7175      ScinaxRuber       60
7176      ScinaxRuber       60
7177      ScinaxRuber       60
7178      ScinaxRuber       60
7179      ScinaxRuber       60
7180      ScinaxRuber       60
7181      ScinaxRuber       60
7182      ScinaxRuber       60
7183      ScinaxRuber       60
7184      ScinaxRuber       60
7185      ScinaxRuber       60
7186      ScinaxRuber       60
7187      ScinaxRuber       60
7188      ScinaxRuber       60
7189      ScinaxRuber       60
7190      ScinaxRuber       60
7191      ScinaxRuber       60
7192      ScinaxRuber       60
7193      ScinaxRuber       60
7194      ScinaxRuber       60

[7195 rows x 26 columns]
```

```
In [2]: X = df.loc[:, :'MFCCs_22'] # independent variables
        labels = df.loc[:, 'Family':]   # lables: Family, Genus, and Species

        # Choose 70% of the data randomly as the training set
        X_train, X_test, labels_train, labels_test = train_test_split(X, labels, test_size=0.3

        # 1. Family
        y1_train = labels_train.loc[:,'Family']
        y1_test = labels_test.loc[:,'Family']
```

```
# 2. Genus
y2_train = labels_train.loc[:,'Genus']
y2_test = labels_test.loc[:,'Genus']

# 3. Species
y3_train = labels_train.loc[:,'Species']
y3_test = labels_test.loc[:,'Species']
```

1. (b) Each instance has three labels: Families, Genus, and Species. Each of the labels has multiple classes. We wish to solve a multi-class and multi-label problem. One of the most important approaches to multi-class classication is to train a classier for each label. We rst try this approach:

i. Research exact match and hamming score/ loss methods for evaluating multilabel classication and use them in evaluating the classiers in this problem.

(Answer) Exact match is the percentage of instances having all their labels classified correctly.
Hamming score is the fraction of correctly classifed labels to the total number of labels.
Hamming loss is the fraction of wrongly classified labels to the total number of labels. Thus, hamming score is '1-Hamming loss.'

1. (b) ii. Train a SVM for each of the labels, using Gaussian kernels and one versus all classiers. Determine the weight of the SVM penalty and the width of the Gaussian Kernel using 10 fold cross validation.1 You are welcome to try to solve the problem with both standardized 2 and raw attributes and report the results.

In [3]:
```
from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV
import numpy as np

# decision_function_shape='ovr': one vesus rest
# kernel='rbf': Gaussian kernel
clf = SVC(kernel='rbf', decision_function_shape='ovr')

# C: the weight of the SVM penalty
# gamma: the width of the Gaussian Kernel
C_range = np.logspace(-2, 2, 5) # 0.01, 0.1, 1, 10, 100
gamma_range = np.logspace(-2, 2, 5)
param_grid = dict(gamma=gamma_range, C=C_range)
# 10 fold cross validation
grid = GridSearchCV(clf, param_grid=param_grid, cv=10)
```

In [4]:
```
# 1. Family (raw data)
# Determine C and gamma by 10 fold CV
grid.fit(X_train, y1_train)
print(grid.best_params_, grid.best_score_)
```

```
{'C': 10.0, 'gamma': 1.0} 0.9912629070691025
```

In [5]:
```
C=grid.best_params_['C']
gamma=grid.best_params_['gamma']
clf = SVC(C=C, gamma=gamma, kernel='rbf', decision_function_shape='ovr')
clf.fit(X_train, y1_train)
print("train accuaracy: %0.4f" % clf.score(X_train, y1_train))
print("test accuaracy: %0.4f" % clf.score(X_test, y1_test))
```

```
train accuaracy: 0.9986
test accuaracy: 0.9893
```

In [6]:
```
from sklearn.metrics import hamming_loss

y1_pred = clf.predict(X_test)
h1 = hamming_loss(y1_test, y1_pred)
print(h1)  # the fraction of wrong labels for Family
```

```
0.010653080129689671
```

In [7]:
```
# 2. Genus (raw data)
# Determine C and gamma by 10 fold CV
grid.fit(X_train, y2_train)
print(grid.best_params_, grid.best_score_)
```

```
{'C': 10.0, 'gamma': 1.0} 0.9892772041302621
```

In [8]:
```
C=grid.best_params_['C']
gamma=grid.best_params_['gamma']
clf = SVC(C=C, gamma=gamma, kernel='rbf', decision_function_shape='ovr')
clf.fit(X_train, y2_train)
print("train accuaracy: %0.4f" % clf.score(X_train, y2_train))
print("test accuaracy: %0.4f" % clf.score(X_test, y2_test))
```

```
train accuaracy: 0.9990
test accuaracy: 0.9889
```

In [9]:
```
y2_pred = clf.predict(X_test)
h2 = hamming_loss(y2_test, y2_pred)
print(h2) # the fraction of wrong labels for Genus
```

```
0.0111162575266327
```

In [10]:
```
# 3. Species (raw data)
# Determine C and gamma by 10 fold CV
grid.fit(X_train, y3_train)
print(grid.best_params_, grid.best_score_)
```

```
{'C': 10.0, 'gamma': 1.0} 0.9882843526608419
```

```
In [11]: C=grid.best_params_['C']
         gamma=grid.best_params_['gamma']
         clf = SVC(C=C, gamma=gamma,kernel='rbf', decision_function_shape='ovr')
         clf.fit(X_train, y3_train)
         print("train accuaracy: %0.4f" % clf.score(X_train, y3_train))
         print("test accuaracy: %0.4f" % clf.score(X_test, y3_test))
```

```
train accuaracy: 0.9988
test accuaracy: 0.9893
```

```
In [12]: y3_pred = clf.predict(X_test)
         h3 = hamming_loss(y3_test, y3_pred)
         print(h3) # the fraction of wrong labels for Species
```

```
0.010653080129689671
```

```
In [13]: # exact match
         # the percentage of instances having all their labels classified correctly

         count = 0
         for i in range(len(y1_test)):
             if y1_pred[i]==y1_test.iloc[i] and \
             y2_pred[i]==y2_test.iloc[i] and \
             y3_pred[i]==y3_test.iloc[i]:
                 count += 1

         e =(count*100)/len(y1_test) # exact math
         h=(h1+h2+h3)/3 # hamming loss

         print("exact match: %0.3f" % e, "%")
         print("hamming loss: %0.3f" % h)
```

```
exact match: 98.518 %
hamming loss: 0.011
```

```
In [14]: # standardization
         from sklearn.preprocessing import StandardScaler

         scX_train = StandardScaler().fit_transform(X_train)
         scX_test = StandardScaler().fit_transform(X_test)
```

```
In [15]: # 1. Family (standardized data)
         # Determine C and gamma by 10 fold CV
         grid.fit(scX_train, y1_train)
         print(grid.best_params_, grid.best_score_)
```

```
{'C': 10.0, 'gamma': 0.1} 0.9900714853057982
```

```
In [17]: C=grid.best_params_['C']
         gamma=grid.best_params_['gamma']
         clf = SVC(C=C, gamma=gamma, kernel='rbf', decision_function_shape='ovr')
         clf.fit(scX_train, y1_train)
         print("train accuaracy: %0.4f" % clf.score(scX_train, y1_train))
         print("test accuaracy: %0.4f" % clf.score(scX_test, y1_test))
```

```
train accuaracy: 1.0000
test accuaracy: 0.9903
```

```
In [18]: y1_pred = clf.predict(scX_test)
         h1 = hamming_loss(y1_test, y1_pred)
         print(h1)  # the fraction of wrong labels for Family
```

```
0.009726725335803613
```

```
In [19]: # 2. Genus (standardized data)
         # Determine C and gamma by 10 fold CV
         grid.fit(scX_train, y2_train)
         print(grid.best_params_, grid.best_score_)
```

```
{'C': 10.0, 'gamma': 0.01} 0.988482922954726
```

```
In [20]: C=grid.best_params_['C']
         gamma=grid.best_params_['gamma']
         clf = SVC(C=C, gamma=gamma, kernel='rbf', decision_function_shape='ovr')
         clf.fit(scX_train, y2_train)
         print("train accuaracy: %0.4f" % clf.score(scX_train, y2_train))
         print("test accuaracy: %0.4f" % clf.score(scX_test, y2_test))
```

```
train accuaracy: 0.9958
test accuaracy: 0.9852
```

```
In [21]: y2_pred = clf.predict(scX_test)
         h2 = hamming_loss(y2_test, y2_pred)
         print(h2)  # the fraction of wrong labels for Genus
```

```
0.014821676702176934
```

```
In [22]: # 3. Species (standardized data)
         # Determine C and gamma by 10 fold CV
         grid.fit(scX_train, y3_train)
         print(grid.best_params_, grid.best_score_)
```

```
{'C': 10.0, 'gamma': 0.01} 0.9882843526608419
```

```
In [24]: C=grid.best_params_['C']
         gamma=grid.best_params_['gamma']
         clf = SVC(C=C, gamma=gamma, kernel='rbf', decision_function_shape='ovr')
         clf.fit(scX_train, y3_train)
         print("train accuaracy: %0.4f" % clf.score(scX_train, y3_train))
         print("test accuaracy: %0.4f" % clf.score(scX_test, y3_test))
```

```
train accuaracy: 0.9968
test accuaracy: 0.9889
```

```
In [25]: y3_pred = clf.predict(scX_test)
         h3 = hamming_loss(y3_test, y3_pred)
         print(h3)  # the fraction of wrong labels for Species
```

```
0.0111162575266327
```

```
In [26]: # exact match

         count = 0
         for i in range(len(y1_test)):
             if y1_pred[i]==y1_test.iloc[i] and \
             y2_pred[i]==y2_test.iloc[i] and \
             y3_pred[i]==y3_test.iloc[i]:
                 count += 1

         e =(count*100)/len(y1_test) # exact math
         h=(h1+h2+h3)/3 # hamming loss

         print("exact match: %0.3f" % e, "%")
         print("hamming loss: %0.3f" % h)
```

```
exact match: 97.869 %
hamming loss: 0.012
```

(Answer of 1. (b) ii) 1. raw data
exact match: 98.518 %
hamming loss: 0.011

2. standardized data

exact match: 97.869 %
hamming loss: 0.012
The performance of the raw data is sligthly better than that of the standardized data.

1. (b) iii. Repeat 1(b)ii with L1-penalized SVMs.3 Remember to standardize4 the attributes. Determine the weight of the SVM penalty using 10 fold cross validation.

```
In [27]: from sklearn.svm import LinearSVC

         clf = LinearSVC(penalty='l1', multi_class='ovr', dual=False)
         C_range = np.logspace(-2, 5, 8) # 10^-2, 10^-1, 1, 10, ...10^5
         param_grid = dict(C=C_range)

         grid = GridSearchCV(clf, param_grid=param_grid, cv=10)

In [28]: # 1. Family (L1-penalized SVM)
         # Determine C by 10 fold CV
         grid.fit(scX_train, y1_train)
         print(grid.best_params_, grid.best_score_)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
```

```
{'C': 100.0} 0.9418189038919778
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
```

```
  "the number of iterations.", ConvergenceWarning)
```

```
In [29]: clf = LinearSVC(C=grid.best_params_['C'], penalty='l1', multi_class='ovr', dual=False
         clf.fit(scX_train, y1_train)
         print("train accuaracy: %0.4f" % clf.score(scX_train, y1_train))
         print("test accuaracy: %0.4f" % clf.score(scX_test, y1_test))
```

```
train accuaracy: 0.9432
test accuaracy: 0.9264


C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
In [30]: y1_pred = clf.predict(scX_test)
         h1 = hamming_loss(y1_test, y1_pred)
         print(h1)
```

```
0.07364520611394164
```

```
In [31]: # 2. Genus (L1-penalized SVM)
         # Determine C by 10 fold CV
         grid.fit(scX_train, y2_train)
         print(grid.best_params_, grid.best_score_)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
```

```
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
```

```
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
```

```
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
{'C': 10.0} 0.9547259729944401
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

In [32]:
```python
clf = LinearSVC(C=grid.best_params_['C'], penalty='l1', multi_class='ovr', dual=False
clf.fit(scX_train, y2_train)
print("train accuaracy: %0.4f" % clf.score(scX_train, y2_train))
print("test accuaracy: %0.4f" % clf.score(scX_test, y2_test))
```

```
train accuaracy: 0.9585
test accuaracy: 0.9416
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

In [33]:
```python
y2_pred = clf.predict(scX_test)
h2 = hamming_loss(y2_test, y2_pred)
print(h2)
```

```
0.058360352014821676
```

In [34]:
```python
# 3. Species (L1-penalized SVM)
# Determine C by 10 fold CV
grid.fit(scX_train, y3_train)
print(grid.best_params_, grid.best_score_)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
{'C': 1.0} 0.9618745035742653
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
In [35]: clf = LinearSVC(C=grid.best_params_['C'], penalty='l1', multi_class='ovr', dual=False
         clf.fit(scX_train, y3_train)
         print("train accuaracy: %0.4f" % clf.score(scX_train, y3_train))
         print("test accuaracy: %0.4f" % clf.score(scX_test, y3_test))
```

```
train accuaracy: 0.9658
test accuaracy: 0.9467
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
In [36]: y3_pred = clf.predict(scX_test)
         h3 = hamming_loss(y3_test, y3_pred)
         print(h3)
```

```
0.053265400648448355
```

In [37]:
```
# exact match

count = 0
for i in range(len(y1_test)):
    if y1_pred[i]==y1_test.iloc[i] and \
    y2_pred[i]==y2_test.iloc[i] and \
    y3_pred[i]==y3_test.iloc[i]:
        count += 1

e =(count*100)/len(y1_test) # exact math
h=(h1+h2+h3)/3 # hamming loss

print("exact match: %0.3f" % e, "%")
print("hamming loss: %0.3f" % h)
```

```
exact match: 89.764 %
hamming loss: 0.062
```

1. (b) iv. Repeat 1(b)iii by using SMOTE or any other method you know to remedy class imbalance. Report your conclusions about the classiers you trained.

In [38]:
```
# making string classes to numerical classes for smote
from sklearn.preprocessing import LabelEncoder

# converting training label
lb1 = LabelEncoder().fit_transform(y1_train)
lb2 = LabelEncoder().fit_transform(y2_train)
lb3 = LabelEncoder().fit_transform(y3_train)
# converting test label
te1 = LabelEncoder().fit_transform(y1_test)
te2 = LabelEncoder().fit_transform(y2_test)
te3 = LabelEncoder().fit_transform(y3_test)
```

In [39]:
```
# smote
from imblearn.over_sampling import SMOTE

X1_smote, y1_smote = SMOTE().fit_resample(scX_train, lb1)
X2_smote, y2_smote = SMOTE().fit_resample(scX_train, lb2)
X3_smote, y3_smote = SMOTE().fit_resample(scX_train, lb3)
```

In [40]:
```
# 1. Family (L1-penalized SVM)
# Determine C by 10 fold CV
grid.fit(X1_smote, y1_smote)
print(grid.best_params_, grid.best_score_)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
```

```
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
```

```
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
```

```
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
{'C': 10000.0} 0.9524699382515437
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
In [41]: clf = LinearSVC(C=grid.best_params_['C'], penalty='l1', multi_class='ovr', dual=False
         clf.fit(X1_smote, y1_smote)
         print("train accuaracy: %0.4f" % clf.score(X1_smote, y1_smote))
         print("test accuaracy: %0.4f" % clf.score(scX_test, te1))
```

```
train accuaracy: 0.9532
test accuaracy: 0.9097
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
In [42]: y1_pred = clf.predict(scX_test)
         h1 = hamming_loss(te1, y1_pred)
         print(h1)
```

```
0.09031959240389069
```

```
In [43]: # 2. Genus (L1-penalized SVM)
         # Determine C by 10 fold CV
         grid.fit(X2_smote, y2_smote)
         print(grid.best_params_, grid.best_score_)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
{'C': 100.0} 0.9573676928398478
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

In [44]:
```python
clf = LinearSVC(C=grid.best_params_['C'], penalty='l1', multi_class='ovr', dual=False
clf.fit(X2_smote, y2_smote)
print("train accuaracy: %0.4f" % clf.score(X2_smote, y2_smote))
print("test accuaracy: %0.4f" % clf.score(scX_test, te2))
```

```
train accuaracy: 0.9584
test accuaracy: 0.9078
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

In [45]:
```python
y2_pred = clf.predict(scX_test)
h2 = hamming_loss(te2, y2_pred)
print(h2)
```

```
0.0921723019916628
```

```
In [46]: # 3. Species (L1-penalized SVM)
         # Determine C by 10 fold CV
         grid.fit(X3_smote, y3_smote)
         print(grid.best_params_, grid.best_score_)
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
```

```
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Libl
```

```
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
```

```
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

```
{'C': 10.0} 0.9617502054231717
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

In [48]:
```
clf = LinearSVC(C=grid.best_params_['C'], penalty='l1', multi_class='ovr', dual=False
clf.fit(X3_smote, y3_smote)
print("train accuaracy: %0.4f" % clf.score(X3_smote, y3_smote))
print("test accuaracy: %0.4f" % clf.score(scX_test, te3))
```

```
train accuaracy: 0.9625
test accuaracy: 0.9546
```

```
C:\Users\Myunghee\Anaconda3\lib\site-packages\sklearn\svm\base.py:922: ConvergenceWarning: Lib
  "the number of iterations.", ConvergenceWarning)
```

In [49]:
```
y3_pred = clf.predict(scX_test)
h3 = hamming_loss(te3, y3_pred)
print(h3)
```

```
0.04539138490041686
```

In [50]:
```
# exact match

count = 0
for i in range(len(y1_test)):
    if y1_pred[i]==te1[i] and \
    y2_pred[i]==te2[i] and \
    y3_pred[i]==te3[i]:
        count += 1

h=(h1+h2+h3)/3
e =(count*100)/len(y1_test)
print("hamming loss: %0.3f" % h)
print("exact match: %0.3f" % e, "%")
```

```
hamming loss: 0.076
exact match: 85.410 %
```

2. K-Means Clustering on a Multi-Class and Multi-Label Data Set Monte-Carlo Simulation:

Perform the following procedures 50 times, and report the average and standard deviation of the 50 Hamming Distances that you calculate.

(a) Use k-means clustering on the whole Anuran Calls (MFCCs) Data Set (do not split the data into train and test, as we are not performing supervised learning in this exercise). Choose k {1,2,...,50} automatically based on one of the methods provided in the slides (CH or Gap Statistics or scree plots or Silhouettes) or any other method you know.

```
In [51]: from sklearn.cluster import KMeans
         from sklearn.metrics import silhouette_score
         # calinski_harabaz_score (CH index)

         # selecting K among {1, 2, ..., 50}
         def Kmeans_k (X):
             dic = dict()
             for i in range(49):
                 k = i+2
                 kmeans = KMeans(n_clusters=k).fit(X)
                 labels= kmeans.labels_

                 s_avg = silhouette_score(X, labels)
                 dic[k]=s_avg
             k_select=pd.DataFrame(dic, index=['s_avg']).T
             k_select = k_select.sort_values(by=['s_avg'], ascending = False)
             K = k_select.iloc[0].name
             return K
```

```
In [53]: K = Kmeans_k(X)
         print(K)
```

```
4
```

2. (b) In each cluster, determine which family is the majority by reading the true labels. Repeat for genus and species.

(c) Now for each cluster you have a majority label triplet (family, genus, species). Calculate the average Hamming distance, Hamming score, and Hamming loss5 between the true labels and the labels assigned by clusters.

```
In [54]: # making string classes to numerical classes
         family = df.loc[:, 'Family']
         genus = df.loc[:, 'Genus']
```

```
        species = df.loc[:, 'Species']
        l1=LabelEncoder().fit(family)
        l2=LabelEncoder().fit(genus)
        l3=LabelEncoder().fit(species)
        label1 = l1.transform(family)
        label2 = l2.transform(genus)
        label3 = l3.transform(species)
```

```
In [57]: tot_table=dict()

         # repeating 50 times K-means clustering (K=4)
         for i in range(50):
             kmeans = KMeans(n_clusters=K).fit(X)
             labels= kmeans.labels_
             cl_table = pd.DataFrame(labels, columns=['cluster'])
             cl_table.insert(1,"family", label1)
             cl_table.insert(2,"genus", label2)
             cl_table.insert(3,"species", label3)
             cl_table = cl_table.sort_values(by=['cluster'])

             clust=cl_table.iloc[:,0]
             c = np.bincount(clust) # counting # of each cluster

             # splitting data according to clusters
             clust0=cl_table.iloc[:c[0],:]  # cluster 0
             clust1=cl_table.iloc[c[0]:c[0]+c[1],:] # cluster 1
             clust2=cl_table.iloc[c[0]+c[1]:c[0]+c[1]+c[2],:] # cluster 2
             clust3=cl_table.iloc[c[0]+c[1]+c[2]:,:] # cluster 3

             dic_class=dict()
             dic_loss=dict()

             # each cluster with a majority label triplet(family, genus, species)
             for j in range(3):
                 c0=np.bincount(clust0.iloc[:,j+1])
                 c1=np.bincount(clust1.iloc[:,j+1])
                 c2=np.bincount(clust2.iloc[:,j+1])
                 c3=np.bincount(clust3.iloc[:,j+1])
                 m0=np.argmax(c0) # majority label of cluster 0
                 m1=np.argmax(c1) # majority label of cluster 1
                 m2=np.argmax(c2) # majority label of cluster 2
                 m3=np.argmax(c3) # majority label of cluster 3

                 dic_class[j+1]=m0, m1, m2, m3
                 # the # of woringly assigned labels
                 dic_loss[j+1]=len(clust)-(c0[m0]+c1[m1]+c2[m2]+c3[m3])

             # assigned family labels list for each cluster
```

```
        fam=l1.inverse_transform(dic_class[1])
        # assigned genus labels list for each cluster
        gen=l2.inverse_transform(dic_class[2])
        # assigned species labels list for each cluster
        spe=l3.inverse_transform(dic_class[3])

        # hamming loss
        HL=(dic_loss[1]+dic_loss[2]+dic_loss[3])/(len(clust)*3)

        # each 50 iteration, family, genus, species labels for each cluster
        # and hamming loss
        tot_table[i]=fam, gen, spe, HL
```

In [64]:
```
# answers of 2. (b) and (c)
T=pd.DataFrame(tot_table)
print(T)
```

```
                                                 0   \
0  [Hylidae, Leptodactylidae, Dendrobatidae, Hyli...
1        [Hypsiboas, Adenomera, Ameerega, Hypsiboas]
2  [HypsiboasCinerascens, AdenomeraHylaedactylus,...
3                                           0.222423

                                                 1   \
0  [Dendrobatidae, Hylidae, Hylidae, Leptodactyli...
1        [Ameerega, Hypsiboas, Hypsiboas, Adenomera]
2  [Ameeregatrivittata, HypsiboasCinerascens, Hyp...
3                                           0.222423

                                                 2   \
0  [Hylidae, Leptodactylidae, Dendrobatidae, Hyli...
1        [Hypsiboas, Adenomera, Ameerega, Hypsiboas]
2  [HypsiboasCinerascens, AdenomeraHylaedactylus,...
3                                           0.222423

                                                 3   \
0  [Leptodactylidae, Dendrobatidae, Hylidae, Hyli...
1        [Adenomera, Ameerega, Hypsiboas, Hypsiboas]
2  [AdenomeraHylaedactylus, Ameeregatrivittata, H...
3                                           0.221913

                                                 4   \
0  [Leptodactylidae, Hylidae, Dendrobatidae, Hyli...
1        [Adenomera, Hypsiboas, Ameerega, Hypsiboas]
2  [AdenomeraHylaedactylus, HypsiboasCordobae, Am...
3                                           0.221774

                                                 5   \
```

```
0  [Leptodactylidae, Hylidae, Dendrobatidae, Hyli...
1        [Adenomera, Hypsiboas, Ameerega, Hypsiboas]
2  [AdenomeraHylaedactylus, HypsiboasCinerascens,...
3                                           0.222423

                                                   6  \
0  [Leptodactylidae, Hylidae, Dendrobatidae, Hyli...
1        [Adenomera, Hypsiboas, Ameerega, Hypsiboas]
2  [AdenomeraHylaedactylus, HypsiboasCordobae, Am...
3                                           0.222423

                                                   7  \
0  [Leptodactylidae, Dendrobatidae, Hylidae, Hyli...
1        [Adenomera, Ameerega, Hypsiboas, Hypsiboas]
2  [AdenomeraHylaedactylus, Ameeregatrivittata, H...
3                                           0.222423

                                                   8  \
0  [Leptodactylidae, Dendrobatidae, Hylidae, Hyli...
1        [Adenomera, Ameerega, Hypsiboas, Hypsiboas]
2  [AdenomeraHylaedactylus, Ameeregatrivittata, H...
3                                           0.222423

                                                   9  ...  \
0  [Hylidae, Leptodactylidae, Dendrobatidae, Hyli...  ...
1        [Hypsiboas, Adenomera, Ameerega, Hypsiboas]  ...
2  [HypsiboasCinerascens, AdenomeraHylaedactylus,...  ...
3                                           0.222423  ...

                                                  40  \
0  [Hylidae, Leptodactylidae, Hylidae, Dendrobati...
1        [Hypsiboas, Adenomera, Hypsiboas, Ameerega]
2  [HypsiboasCordobae, AdenomeraHylaedactylus, Hy...
3                                           0.222423

                                                  41  \
0  [Dendrobatidae, Hylidae, Leptodactylidae, Hyli...
1        [Ameerega, Hypsiboas, Adenomera, Hypsiboas]
2  [Ameeregatrivittata, HypsiboasCordobae, Adenom...
3                                           0.222423

                                                  42  \
0  [Hylidae, Leptodactylidae, Leptodactylidae, Hy...
1       [Hypsiboas, Adenomera, Adenomera, Hypsiboas]
2  [HypsiboasCordobae, AdenomeraHylaedactylus, Ad...
3                                           0.245263

                                                  43  \
```

```
0  [Leptodactylidae, Hylidae, Hylidae, Dendrobati...
1        [Adenomera, Hypsiboas, Hypsiboas, Ameerega]
2  [AdenomeraHylaedactylus, HypsiboasCinerascens,...
3                                           0.222284

                                                  44  \
0  [Dendrobatidae, Hylidae, Leptodactylidae, Hyli...
1        [Ameerega, Hypsiboas, Adenomera, Hypsiboas]
2  [Ameeregatrivittata, HypsiboasCordobae, Adenom...
3                                           0.222423

                                                  45  \
0  [Leptodactylidae, Hylidae, Leptodactylidae, Hy...
1       [Adenomera, Hypsiboas, Adenomera, Hypsiboas]
2  [AdenomeraHylaedactylus, HypsiboasCordobae, Ad...
3                                           0.245263

                                                  46  \
0  [Dendrobatidae, Hylidae, Hylidae, Leptodactyli...
1        [Ameerega, Hypsiboas, Hypsiboas, Adenomera]
2  [Ameeregatrivittata, HypsiboasCordobae, Hypsib...
3                                           0.222423

                                                  47  \
0  [Leptodactylidae, Leptodactylidae, Hylidae, Hy...
1       [Adenomera, Adenomera, Hypsiboas, Hypsiboas]
2  [AdenomeraAndre, AdenomeraHylaedactylus, Hypsi...
3                                           0.233727

                                                  48  \
0  [Leptodactylidae, Hylidae, Hylidae, Dendrobati...
1        [Adenomera, Hypsiboas, Hypsiboas, Ameerega]
2  [AdenomeraHylaedactylus, HypsiboasCinerascens,...
3                                           0.222284

                                                  49
0  [Hylidae, Leptodactylidae, Dendrobatidae, Hyli...
1        [Hypsiboas, Adenomera, Ameerega, Hypsiboas]
2  [HypsiboasCordobae, AdenomeraHylaedactylus, Am...
3                                           0.222423

[4 rows x 50 columns]
```

Report the average and standard deviation of the 50 Hamming Distances that you calculate.

(Answer) Hamming score is the fraction of correctly classifed labels to the total number of labels.

Hamming loss is the fraction of wrongly classified labels to the total number of labels. Thus, hamming score is '1-Hamming loss.'

Hamming distance is the fraction of wrongly classified labels to the total number of samples. Thus, hamming distance is "Hamming loss X 'the number of labels(in this case: 3).'

Hamming score and hamming distance can be easily calculated from hamming loss, so I calculated only hamming loss in this HW.

```
In [65]: import statistics

         m = statistics.mean(T.iloc[3, :])
         s = np.std(T.iloc[3, :])

         print("avearge: %0.3f" % m)
         print("standard deviation: %0.3f" % s)
```

```
avearge: 0.225
standard deviation: 0.007
```