

Logistic Regression. Homework 2

March 20, 2018

0.1 Logistic Regression Part 6

```
In [2]: # Creating X Train
import csv
import numpy
filename = 'XTrain.csv'
raw_data = open(filename, 'rt')
reader = csv.reader(raw_data, delimiter=',', quoting=csv.QUOTE_NONE)
x = list(reader)
Xtrain = numpy.array(x).astype('float')
print(Xtrain.shape)
```

(580, 26048)

```
In [3]: # Creating YTrain
import csv
import numpy
filename2 = 'YTrain.csv'
raw_data2 = open(filename2, 'rt')
reader = csv.reader(raw_data2, delimiter=',', quoting=csv.QUOTE_NONE)
y = list(reader)
Ytrain = numpy.array(y).astype('float')
print(Ytrain.shape)
```

(580, 1)

```
In [4]: #Import Logistic Regression
from sklearn.linear_model import LogisticRegression
logisticRegr = LogisticRegression()
```

1 Running Logistic Regression on Training Data

```
In [5]: logr = logisticRegr.fit(Xtrain, Ytrain.ravel())
```

```
In [6]: #Importing X Test
import csv
```

```

import numpy
filename = 'XTest.csv'
raw_data = open(filename, 'rt')
reader = csv.reader(raw_data, delimiter=',', quoting=csv.QUOTE_NONE)
x = list(reader)
Xtest = numpy.array(x).astype('float')
print(Xtest.shape)

```

(145, 26048)

```

In [7]: #Predicting X Test
        #logr = logisticRegr.predict(Xtest[0:10])

```

2 1. Accuracy Score

```

In [8]: import csv
import numpy
filename = 'YTest.csv'
raw_data = open(filename, 'rt')
reader = csv.reader(raw_data, delimiter=',', quoting=csv.QUOTE_NONE)
x = list(reader)
Ytest = numpy.array(x).astype('float')
#print(Ytest.shape)

predictions = logisticRegr.predict(Xtest)
score = logisticRegr.score(Xtest, Ytest)
print('Accruacy score is', score)

```

Accruacy score is 0.937931034483

```

In [9]: import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import metrics

cm = metrics.confusion_matrix(Ytest, predictions)
print('Confusion Matrix', cm)

```

Confusion Matrix [[98 0]
[9 38]]

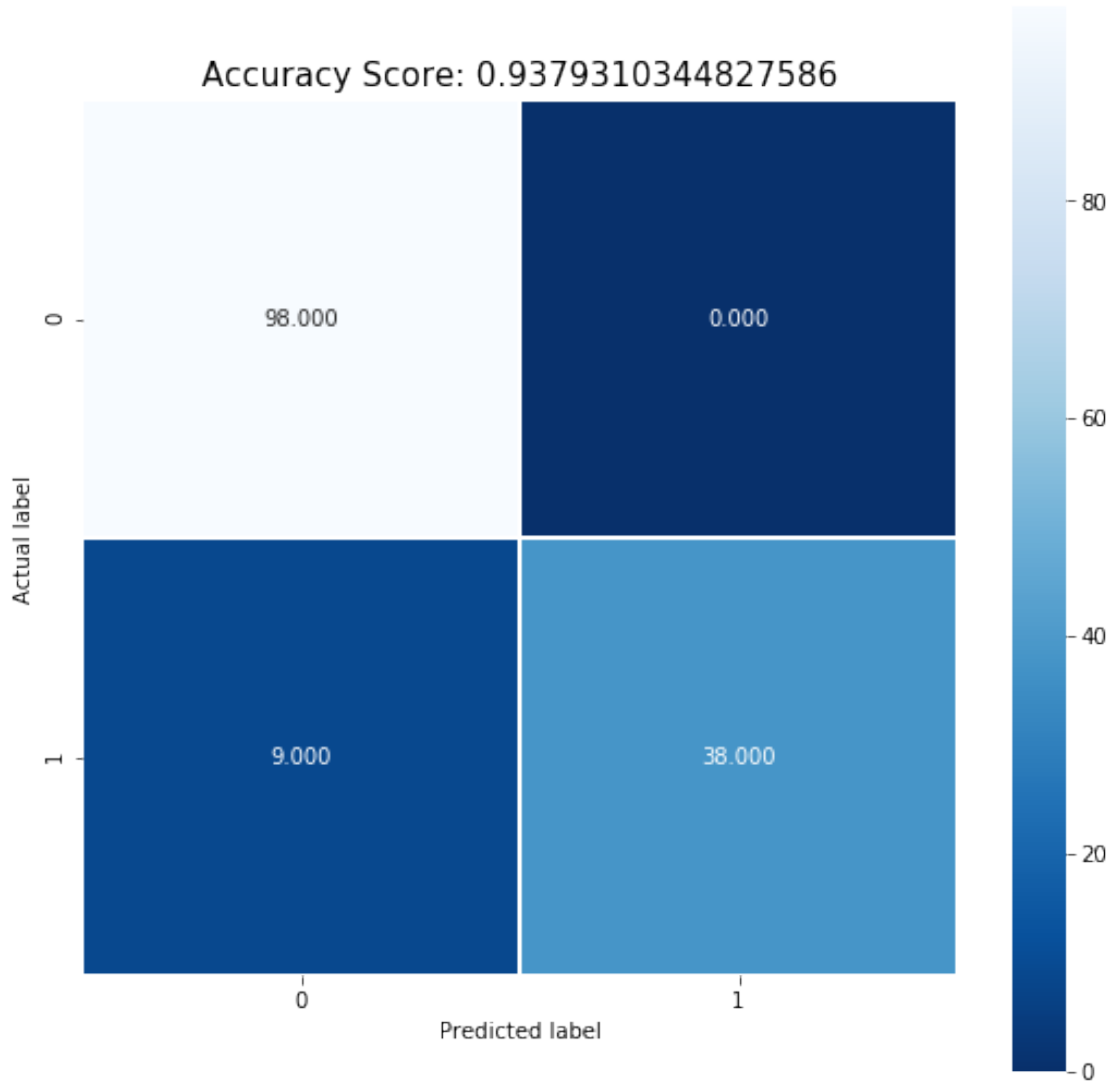
2.1 Confusion Matrix:

```

In [10]: plt.figure(figsize=(9,9))
sns.heatmap(cm, annot=True, fmt=".3f", linewidths=.5, square = True, cmap = 'Blues_r')
plt.ylabel('Actual label');

```

```
plt.xlabel('Predicted label');
all_sample_title = 'Accuracy Score: {0}'.format(score)
plt.title(all_sample_title, size = 15);
plt.show()
```



2.2 2. Plotting Precision and Recall Curve

```
In [11]: from sklearn.metrics import average_precision_score
         average_precision = average_precision_score(Ytest, predictions)

         print('Average precision-recall score: {0:0.2f}'.format(
             average_precision))
```

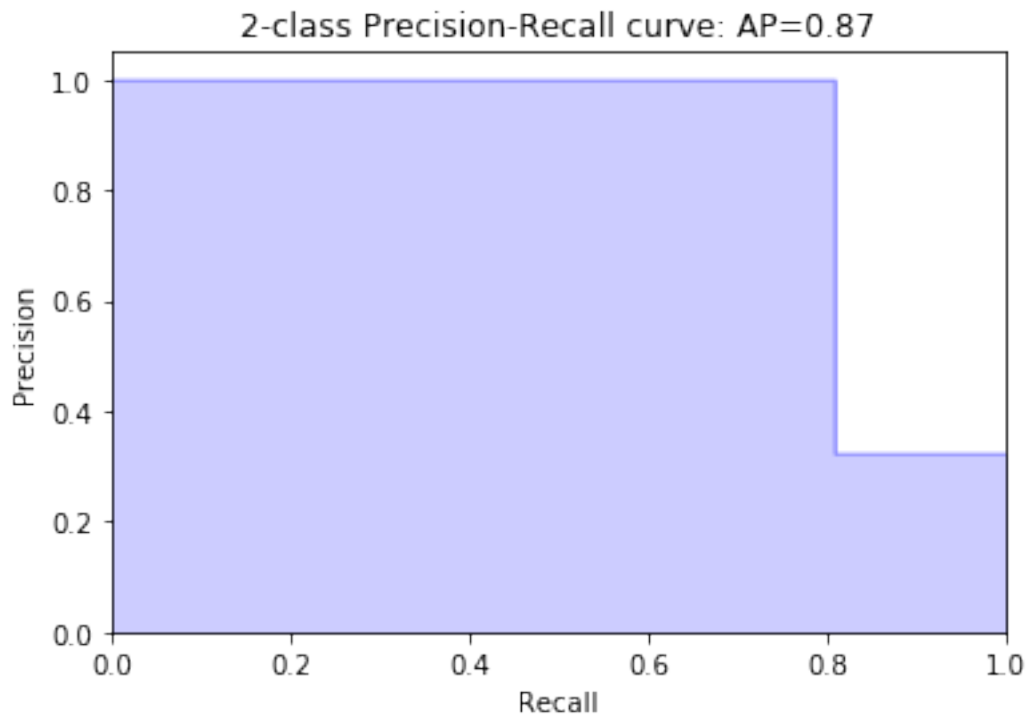
Average precision-recall score: 0.87

```
In [12]: from sklearn.metrics import precision_recall_curve
import matplotlib.pyplot as plt

precision, recall, _ = precision_recall_curve(Ytest, predictions)

plt.step(recall, precision, color='b', alpha=0.2,
        where='post')
plt.fill_between(recall, precision, step='post', alpha=0.2,
                color='b')

plt.xlabel('Recall')
plt.ylabel('Precision')
plt.ylim([0.0, 1.05])
plt.xlim([0.0, 1.0])
plt.title('2-class Precision-Recall curve: AP={0:0.2f}'.format(
        average_precision))
plt.show()
```

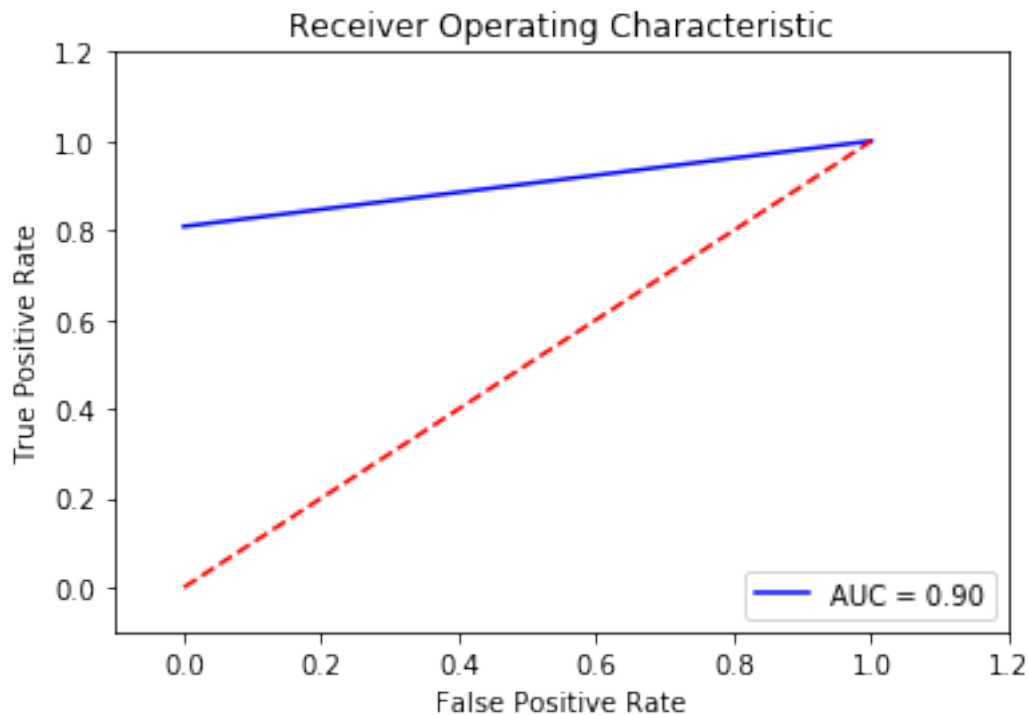


3 Print ROC Curve

```
In [22]: from sklearn.metrics import roc_curve, auc
         from sklearn.preprocessing import label_binarize

false_positive_rate, true_positive_rate, thresholds = roc_curve(Ytest, predictions)
roc_auc = auc(false_positive_rate, true_positive_rate)

plt.title('Receiver Operating Characteristic')
plt.plot(false_positive_rate, true_positive_rate, 'b',
label='AUC = %0.2f'% roc_auc)
plt.legend(loc='lower right')
plt.plot([0,1],[0,1], 'r--')
plt.xlim([-0.1,1.2])
plt.ylim([-0.1,1.2])
plt.ylabel('True Positive Rate')
plt.xlabel('False Positive Rate')
plt.show()
```



3.1 3. Producing Word Cloud From Data

```
In [23]: import numpy
         filename2 = 'vocabulary.csv'
```

```

raw_data2 = open(filename2, 'rt')
reader = csv.reader(raw_data2, delimiter=',', quoting=csv.QUOTE_NONE)
v = list(reader)
vocablist = numpy.array(v).astype('str')

positives = numpy.nonzero(logr.coef_[0] >= 0)
negatives = numpy.nonzero(logr.coef_[0] < 0)
wordslst = {}
for word, coeff in zip(vocablist[:,0], numpy.abs(logr.coef_[0])):
    wordslst[word] = coeff

```

In [24]: wordslst

```

Out[24]: {'after': 0.079414783459440863,
'a': 0.0067958150939156567,
'good': 0.014344651403128537,
'summer': 0.00061921814282962578,
'romano': 0.00082563330524881829,
'prodi': 0.019134911221694813,
'govern': 0.24073417616369466,
'face': 0.031768583766491616,
'harder': 0.022409892621897725,
'econom': 0.074587096174562995,
'time': 0.051650423408707,
'has': 0.083562514621187772,
'there': 0.017433684053606232,
'ever': 0.047799880377020557,
'been': 0.02938571572896901,
'sunnier': 0.019142679344204779,
'italian': 0.025826620627920353,
'august': 0.014097930195211973,
'was': 0.033254499909630274,
'dull': 0.0042998360823884134,
'but': 0.056471241253607272,
'june': 0.0033295607144512144,
'radiant': 0.015177935871408808,
'and': 0.025477224028614194,
'juli': 0.0036949080162317228,
'scorch': 0.013240344623691771,
'septemb': 0.031050963775234256,
'is': 0.044658518580499872,
'prove': 0.029909105312966564,
'unseason': 0.019276784745533744,
'hot': 0.0010709873025075297,
'the': 0.022921829024258204,
'sunni': 0.028313798906981377,
'not': 0.059194432941953638,

```

'confin': 0.0016450260963792429,
'to': 0.013359240459196153,
'weather': 0.0081226695187850608,
'at': 0.01718648231824622,
'first': 0.019972782913783852,
'blush': 0.020139510137061296,
'news': 0.002840045976303089,
'pile': 0.0054276352574008715,
'up': 0.037778749254081333,
'on': 0.025826334019826109,
'desk': 0.011013887538531027,
'of': 0.021957169694380867,
'itali': 0.025392961595001529,
'financ': 0.09303777728336507,
'minist': 0.15920097919955103,
'tommaso': 0.0062806771550384406,
'padoaschioppa': 0.0025735847038951794,
'sinc': 0.021255696985917772,
'he': 0.0045422437321621057,
'said': 0.28962035735686015,
'in': 0.043517869312929371,
'that': 0.010253909736279376,
'state': 0.0090125277736587441,
'public': 0.020471865921640634,
'account': 0.051649483001768734,
'wors': 0.038415308809755536,
'than': 0.019626524690233432,
'1992': 0.015201035853303405,
'year': 0.057013415112858658,
'britain': 0.053398813969134117,
'were': 0.078083705049103377,
'forc': 0.060259712702057343,
'out': 0.075728835968771224,
'europ': 0.052606621834921603,
'exchanger': 0.0044294084075148865,
'mechan': 0.0041019512759199661,
'it': 0.012131709728332274,
'becam': 0.0080205401479792222,
'clear': 0.034023901034664995,
'economi': 0.059674341390667768,
'rebound': 0.02443527782657064,
'faster': 0.051888106597223552,
'expect': 0.033777718030192902,
'last': 0.039496534274501768,
'week': 0.024135724893319808,
'parisbas': 0.0011457404085259753,
'oecd': 0.023968255736865913,
'rais': 0.077886712414638903,

'forecast': 0.038189212525451363,
'2006': 0.0083964851018255657,
'gdp': 0.014511269615007288,
'growth': 0.023261890805900218,
'from': 0.024093520215997314,
'14': 0.0078885595937579229,
'18': 0.0061514618706333297,
'though': 0.039512063678064445,
'this': 0.013956815430146239,
'still': 0.045403092950017883,
'slowest': 0.0083016806348769137,
'then': 0.01800542561404091,
'found': 0.0099421563884582956,
'enhanc': 0.00098984834827050032,
'feed': 0.012180179977118781,
'into': 0.031179900544869743,
'higher': 0.010013155295643946,
'tax': 0.048196191617095908,
'revenu': 0.0029730594704386274,
'so': 0.030298640463680714,
'much': 0.082789281634852524,
'leader': 0.088498724740536661,
'centreleft': 0.027435260750064108,
'abl': 0.064411888595407557,
'announc': 0.079025928822704442,
'jettison': 0.01932547936170571,
'eur5': 0.017181314405636073,
'billion': 0.0075221797637357044,
'65': 0.0036670662060456099,
'spend': 0.014833014416699148,
'cut': 0.0066069416076303383,
'had': 0.03082663929487554,
'plan': 0.017186372553295359,
'for': 0.016238309861883841,
'2007': 0.035545681624812302,
'did': 0.02874739289657359,
'want': 0.039384828877746317,
'make': 0.0024533223919214883,
'weep': 0.023282133004799442,
'some': 0.02378040627969788,
'figur': 0.0051490568925374328,
'have': 0.041527983157967983,
'got': 0.077002479115293093,
'better': 0.004947950554451094,
'borrow': 0.0082285235578195143,
'requir': 0.011857319163688193,
'eight': 0.043390333986796252,
'month': 0.035526906439606265,

'eur364': 0.0040910092573706322,
'eur22': 0.010539146505554745,
'less': 0.10502944821302623,
'2005': 0.062193117318100344,
'offici': 0.0027786314532446178,
'are': 0.087658608024085177,
'now': 0.12774874597459548,
'pencil': 0.033134568720764555,
'budget': 0.014517742595540204,
'deficit': 0.00028506766374718895,
'3839': 0.002091846826251968,
'an': 0.0069591992080431219,
'analysi': 0.027510896043655187,
'just': 0.096757301927590841,
'post': 0.015351047408151177,
'websit': 0.0040798094408423099,
'3wwwlavoceinfo': 0.01133026406802148,
'two': 0.033092401995297283,
'countri': 0.15830962949212732,
'lead': 0.022189258165253696,
'publicfin': 0.0036977487800149656,
'expert': 0.040691500071034274,
'tito': 0.027916381353976987,
'boeri': 0.0073591760670899736,
'pietro': 0.0048948589888579445,
'garibaldi': 0.021740889998226983,
'argu': 0.074250858614244034,
'pessimist': 0.0018225716806106931,
'nobodi': 0.024782893905401764,
'should': 0.018714349848923312,
'be': 0.0028562260604080791,
'surpris': 0.029214217811736647,
'if': 0.0049688284851534632,
'we': 0.28783706004562631,
'end': 0.032093073567911516,
'far': 0.067388194087960526,
'off': 0.03071735150322312,
'euroarea': 0.008427058202794659,
'limit': 0.023541555246836925,
'3': 0.03646788712266609,
'either': 0.011586123825617215,
'way': 0.035070006025872284,
'pictur': 0.0047094670260001466,
'look': 0.0050193050775072193,
'healthier': 0.002418538600220821,
'46': 0.021961621203892851,
'next': 0.045608913654345579,
'by': 0.0027325670931579752,

'see': 0.031129261539447152,
'chart': 0.027330679998630227,
'all': 0.026549979180075671,
'seem': 0.036692575424528306,
'bizarr': 0.0053409637785022694,
'odd': 0.024798776415611668,
'with': 0.01992396416873227,
'grow': 0.010567311220167751,
'belief': 0.041354680643408449,
'particular': 0.076323953840202713,
'london': 0.038676308745339845,
'may': 0.10428398121350543,
'risk': 0.016757890769573146,
'fall': 0.0033159339731600656,
'euro': 0.0056230740853010653,
'studi': 0.027321213106565982,
'publish': 0.047304215629301559,
'short': 0.018256733318069568,
'londonbas': 0.0025383317098002229,
'thinktank': 0.014774440772799879,
'centr': 0.062557478994330548,
'european': 0.0024219845807764095,
'reform': 0.11201781776866024,
'cer': 0.030277468353047794,
'put': 0.039406671954083053,
'happen': 0.038981589179375178,
'daunt': 0.0051601373698699935,
'40': 0.060231790789138033,
'mr': 0.28280797503324751,
'scath': 0.011360314329696591,
'about': 0.0028036529233556795,
'such': 0.027284622443379147,
'talk': 0.079042409128826208,
'my': 0.22391255696129525,
'rather': 0.03587935457113281,
'i': 0.33418310753362113,
'decid': 0.013140634734076983,
'ani': 0.014953258584262373,
'nonsens': 0.0101548713345995,
'scenario': 0.0015763226245694513,
'say': 0.06837869839395061,
'fact': 0.035432255947049476,
'variant': 0.0064798668450764749,
'one': 0.019650762225242705,
'envisag': 0.0068536671332805358,
'deterior': 0.01454108916480676,
'point': 0.070832153527063374,
'where': 0.054624100393066921,

'debt': 0.012905256275050032,
'lose': 0.012791164555593471,
'invest': 0.037652053566693711,
'grade': 0.0070359281981626897,
'which': 0.055458445143305016,
'would': 0.04165210250589145,
'onli': 0.0060107560252229972,
'further': 0.0079797640401589771,
'downgrad': 0.0084230851311486894,
'main': 0.068269768650905949,
'rate': 0.017298677218714652,
'agenc': 0.018227940881755292,
'sharpli': 0.021561245760100881,
'debt serv': 0.0085066265033855552,
'cost': 0.04380959628308987,
'pitch': 0.0067419983597636858,
'downward': 0.0035254856945256918,
'spiral': 0.013959185063137021,
'surg': 0.006482147224192927,
'interest': 0.020066769220712675,
'rise': 0.040410417243354521,
'latest': 0.053837846606477127,
'data': 0.0082367832267047931,
'outcom': 0.0044842761632506375,
'like': 0.0092605559734395635,
'immedi': 0.029691183608868466,
'futur': 0.016505762958221485,
'danger': 0.014980628604431799,
'victim': 5.3414624756265453e-05,
'complac': 0.008139267391588623,
'leftw': 0.0038596593970504742,
'parti': 0.10520105764906092,
'coalit': 0.0097847781125110431,
'will': 0.013370120512434731,
'seiz': 0.033593817980976014,
'report': 0.20217753189984372,
'smaller': 0.031560735586613646,
'demand': 0.00018786859836687319,
'fewer': 0.011323842078572191,
'world': 0.018261410253519041,
'sudden': 0.076444713600609351,
'turn': 0.039336416126793276,
'down': 0.021673112721522465,
'no': 0.051249876075523254,
'mean': 0.079125669623653513,
'imposs': 0.020516103608894682,
'could': 0.060970088320199789,
'substanti': 0.010292288296022096,

'more': 0.037988765706245271,
'tell': 0.051108217490037765,
'thirdlargest': 0.002248677233094314,
'rose': 0.0031010914130975417,
'as': 0.015575389441783711,
'percentag': 0.011006583199476375,
'around': 0.037401418143500127,
'108': 0.010748264014552363,
'primari': 0.059452915948054519,
'surplus': 0.010819794772093973,
'ie': 0.0014361466500555651,
'befor': 0.030063213998875349,
'payment': 0.023628672217888023,
'shrank': 0.021680388831786381,
'05': 0.006347679373972041,
'almost': 0.025632240242714155,
'5': 0.033057552280947804,
'2000': 0.031486820547223536,
'pamphlet': 0.015576249387328048,
'note': 0.017653782649661384,
'need': 0.040257882250869639,
'run': 0.036245808697479766,
'23': 0.034112861942149006,
'prevent': 0.013087244258579801,
'ratio': 0.0069511892891695563,
'other': 0.039820462406306427,
'word': 0.04729238412345637,
'warn': 0.012342876877446635,
'earlier': 0.032294830790754687,
'object': 0.016864077273366463,
'true': 0.011064028634880325,
'recent': 0.028557595941656105,
'improv': 0.057788174883545276,
'under': 0.10972141093318087,
'situat': 0.058897258686635201,
'given': 0.03986205064343127,
'readi': 0.029652562999591337,
'trim': 0.021039966158656491,
'his': 0.011787133493844791,
'caus': 0.041447232340928022,
'concern': 0.019231430773409608,
'disappoint': 0.035580784421571025,
'execut': 0.058183600133840579,
'board': 0.015563153242253137,
'member': 0.0068634620912848086,
'central': 0.058893431192534436,
'bank': 0.010750661611668809,
'lorenzo': 0.0099494196326724554,

'bini': 0.011277096943551031,
'smaghi': 0.012641736390446296,
'increas': 0.053413745769271583,
'temporari': 0.024091482940736642,
'factor': 0.011218276538838076,
'repeat': 0.022715619895438768,
'joaquin': 0.0023954258770367857,
'almunia': 0.016757647266509926,
'commission': 0.031030169518869812,
'stress': 0.0051584380373631685,
'whose': 0.059731499298354064,
'exceed': 0.0013581843949763306,
'annual': 0.024043716133202175,
'output': 0.0047268870925497109,
'tidi': 0.0069812102295577468,
'mere': 0.051028976046568628,
'case': 0.028393580494881498,
'bring': 0.0025278025157042068,
'seen': 0.020778753531541655,
'brussel': 0.028771673529869184,
'big': 0.057518828825568344,
'test': 0.011532832741114319,
'new': 0.013696992150139415,
'weaker': 0.024117142429204642,
'stabil': 0.0092087336568591662,
'pact': 0.02818181807672648,
'suppos': 0.014665021318417543,
'set': 0.013197625435625662,
'fiscal': 0.017116791830404127,
'rule': 0.084373505589082312,
'get': 0.062200034781797885,
'flexibl': 0.0090349655938911921,
'bad': 0.015580005482500754,
'timesbut': 0.0033422549762499587,
'they': 0.0098689001810289573,
'meant': 0.010169718185455557,
'show': 0.036460196502501103,
'greater': 0.02731169703777064,
'disciplin': 0.0047454306817955915,
'cabinet': 0.021042147031680381,
'agre': 0.039959379824013201,
'relax': 0.0090620873968601592,
'becaus': 0.061097681210545866,
'match': 0.0068122428412185883,
'structur': 0.012516858141741433,
'element': 0.014535912585839748,
'detect': 0.016894367473660996,
'accept': 0.026713607244674732,

'basic': 0.0077241105414268801,
'adjust': 0.00643910952913271,
'propos': 0.075033170588420336,
'howev': 0.085214894050287052,
'biggest': 0.039519131722791659,
'communist': 0.048785775280325142,
'refus': 0.0071848881575334823,
'underwrit': 0.013070176745740196,
'mediumterm': 0.010419704048470127,
'strategi': 0.027994323940749563,
'record': 0.050126863894623341,
'unsustain': 0.0078957505885558065,
'aim': 0.060263371108386532,
'pension': 0.024949467488751868,
'health': 0.029963025796555787,
'publicsector': 0.0059312500270274071,
'employ': 0.015723828132723612,
'localgovern': 0.0024987752219832386,
'these': 0.014208319189051346,
'unpopular': 0.038920064702473292,
'or': 0.020376677071564975,
'within': 0.0033643804072033172,
'yet': 0.014871862084505921,
'second': 0.037225674989502341,
'leav': 0.004723096771150904,
'perhap': 0.019976440484772224,
'worri': 0.067217799816833235,
'pull': 0.0068606478141005067,
'competit': 0.01050914751369092,
'erod': 0.00034513246042421433,
'membership': 0.002787564533691294,
'becom': 0.057575368254884962,
'too': 0.054756671320105546,
'pain': 0.01847210000071161,
'inde': 0.057120046970894352,
'continu': 0.089659216987163545,
'loss': 0.023270923856182894,
'problem': 0.11992266962311909,
'conclud': 0.033635995621448408,
'five': 0.025181968153268956,
'minut': 0.033586990806777801,
'midnight': 0.014666324839009724,
'act': 0.018594667470359469,
'ensur': 0.0067439763734238253,
'longterm': 0.0094345061566624318,
'zone': 0.0029139711628356846,
'over': 0.012823604696896524,
'past': 0.034929914728353259,

'few': 0.053682253037034958,
'alarm': 0.0092449622973868502,
'between': 0.050128343852123884,
'2004': 0.024458973342448197,
'total': 0.074658196696659726,
'product': 0.030495397781963644,
'fell': 0.022045363300430183,
'averag': 0.015895450037981814,
'12': 0.00024446023897678579,
'worst': 0.011567423338111124,
'perform': 0.00019255212556708833,
'eu': 0.017354770730155372,
'15': 1.7979611838899964e-05,
'same': 0.014322172946439634,
'both': 0.090316645697556014,
'inflat': 0.024057766963484242,
'nomin': 0.038682920414692602,
'wage': 0.015816090682291884,
'especi': 0.02721568162054629,
'germani': 0.041259268736298912,
'task': 0.035963093945789178,
'fit': 0.012970921600477832,
'depend': 0.001997544990752395,
'larg': 0.010115362542990634,
'effort': 0.034184044307410115,
'privat': 0.050232280365365853,
'sector': 0.0019797770781007887,
'extent': 0.0073380649632423947,
'can': 0.060407491336247315,
'help': 0.10261775694293722,
'liberalis': 0.0024178822282306354,
'ministri': 0.039647891307649188,
'own': 0.11736297541546695,
'defiant': 0.014294112575242757,
'remain': 0.0020459484784499183,
'crucial': 0.0066939805557910516,
'prospect': 0.035940385473413342,
'effect': 0.020584479484492799,
'alreadi': 0.025057753924341813,
'use': 0.039747310511502154,
'unexpect': 0.033515961500392467,
'windfal': 0.010529463570814841,
'three': 0.038263578337495947,
'ahead': 0.023155846950266113,
'without': 0.0067715067476249993,
'elect': 0.10895837305572194,
'when': 0.11755995872056113,
'push': 0.010015305022212741,

'through': 0.038055179186369338,
'fail': 0.018302986893538861,
'do': 0.059443641352300222,
'resist': 0.019469032867097762,
'consequ': 0.003212449485976549,
'dire': 0.006777397086064829,
'warmth': 0.018701484475497642,
'give': 0.016633632891648983,
'colder': 0.022025012740782404,
'quick': 0.021154563065170634,
'preelect': 0.0028865634003562303,
'tension': 0.0025686657064201554,
'among': 0.032545661395276107,
'kurd': 0.0081589630111487468,
'turkey': 0.010286153423187731,
'southeast': 0.0091857581557822539,
'goldplat': 0.017409029463557116,
'pistol': 0.010290252592125985,
'hand': 0.0092114797930707044,
'bead': 0.0084366710750516279,
'hazim': 0.0055238709484523724,
'babat': 0.015325584672983315,
'sit': 0.064356148967777554,
'foot': 0.004447433300330832,
'mountain': 0.022787833913059809,
'rang': 0.0015019161080555839,
'separ': 0.052248716009837784,
'iraq': 0.013178019246399705,
'contempl': 0.0055466551867762327,
'war': 0.018935019385348018,
'chieftain': 0.024385132563645651,
'clan': 0.023948918235053979,
'fight': 0.028534743158122614,
'alongsid': 0.0080811750047241528,
'turkish': 0.0045763442310122453,
'arm': 0.039850590411087107,
'against': 0.060474230675897335,
'kurdistan': 0.012499047627767235,
'worker': 0.027054346972256803,
'pkk': 0.020440980912715664,
'separatist': 0.0095126098969444978,
'guerrilla': 0.008742418542341944,
'hunt': 0.00022749264225668613,
'milit': 0.014815249180302198,
'peshmerga': 0.019405513538986702,
'warrior': 0.028835763390622871,
'democrat': 0.10549989181866261,
'kdp': 0.018501244004343997,

'northern': 0.03819416286840345,
'led': 0.057324759510756759,
'massoud': 0.0021113079225692517,
'barzani': 0.0097765585165504546,
'slaughter': 0.014725936009718117,
'our': 0.19857417032583413,
'women': 0.038689129293633502,
'babi': 0.043351256569161237,
'go': 0.027554145321892747,
'pay': 0.045222304248758588,
'price': 0.040175618028483767,
'vow': 0.0070019339493088122,
'chang': 0.11878615178014167,
'begin': 0.0073408824132907956,
'unit': 0.011562483664474732,
'wherev': 0.0060012517214430095,
'live': 0.020990931504171865,
'nowher': 0.021907966875648381,
'truer': 0.018769132243402865,
'american': 0.0039062782147397546,
'support': 0.015645549597453907,
'mani': 0.0038120978322124018,
'4m': 6.7174210151454402e-05,
'come': 0.046479349075619548,
'closer': 0.0192837296618735,
'achiev': 0.018304945880604693,
'fullblown': 0.0055535071950781013,
'independ': 0.058145786213760681,
'who': 0.0064060672374387914,
'kurdishcontrol': 0.0023109582303997697,
'enclav': 0.015515354720416253,
'declar': 0.069204110024834647,
'day': 0.043703377392873471,
'kurdish': 0.014091122233735457,
'fratricid': 0.014945621574171796,
'let': 0.0050129264602500143,
'soldier': 0.0018347911419531386,
'overrun': 0.011985074557776457,
'territori': 0.0029315680728707288,
'order': 0.0075882483082847811,
'attack': 0.052965591056005193,
'real': 0.0037064422329666281,
'opin': 0.0048090293235127386,
'exist': 0.017773583321712291,
'despit': 0.06453745630268716,
'call': 0.10104997000436491,
'crossbord': 0.00019207696696808967,
'action': 0.0039047208323889698,

'armi': 0.032060195188437814,
'target': 0.025488168151650722,
'quasiindepend': 0.0065206826261957011,
'general': 0.017158577041057277,
'existenti': 0.0104796555095929,
'threat': 0.014268928542782545,
'stoke': 0.0015438597250823903,
'passion': 0.0018473462162849276,
'14modd': 0.0036411470398978289,
'emerg': 0.0067018258788687102,
'matter': 0.028454636153034046,
'assert': 0.026985430224715085,
'ibrahim': 0.013393135005041542,
'guclu': 0.010499218317155364,
'veteran': 0.0033082326455255583,
'politician': 0.091985055714022532,
'togeth': 0.021311908843462048,
'drive': 0.0095450816937217731,
'iraqi': 0.0061721346755136408,
'add': 0.013121069035165142,
'mind': 0.043598716782068406,
'america': 0.057382726831178593,
'opposit': 0.080882162280257769,
'vote': 0.058662611243417734,
'mild': 0.0087080176068860798,
'islamist': 0.035987725691999838,
'prime': 0.059294336575164586,
'recep': 0.00019586482992693489,
'tayyip': 0.0058791756164870918,
'erdogan': 0.0026567532369980978,
'incurs': 0.018648024369720757,
'least': 0.0021743276913385409,
'22nd': 0.018685273724501791,
'mass': 0.041542979864579256,
'along': 0.020249929722271031,
'border': 0.027808726999539558,
'beefedup': 0.030150826837506604,
'presenc': 0.01104917398929851,
'palpabl': 0.0061110092566425481,
'sirnak': 0.0091537492582429037,
'provinc': 0.025777774163669033,
'nogo': 0.032029076954722541,
'secur': 0.010848790448073839,
'armour': 0.0009140084643667316,
'personnelcarri': 0.0027815044495970056,
'gun': 0.0025721683076458573,
'passersbi': 0.019953310427549997,
'helicopt': 0.023560508567258324,

'clatter': 0.0069013044261725517,
'overhead': 0.00079306907659707122,
'region': 0.02668884705452533,
'capit': 0.038556734100556093,
'diyarbakir': 0.00016635208903587466,
'your': 0.08760728528821897,
'correspond': 0.01396756620148811,
'count': 0.027656796025459757,
'nine': 0.016581799974330718,
'f16': 0.0097746839296754956,
'fighter': 0.0022308737826067216,
'jet': 0.024177602935402158,
'screech': 0.021553903612983316,
'toward': 0.043137400392969247,
'space': 0.0080916827629088639,
'20': 0.015464978532396024,
'hit': 0.030121102316386164,
'back': 0.0078299258187790374,
'murder': 0.025587295548994308,
'civilian': 0.015879691238450549,
'alik': 0.011278589822097727,
'ayla': 0.0097877820851699857,
'akat': 0.012949345053086907,
'humanright': 0.0088901966366211008,
'lawyer': 0.0049535506775522592,
'stand': 0.00026983101990177776,
'batman': 0.0045448208410999689,
'prokurdish': 0.0056521465645274962,
'dtp': 0.0058012638472359065,
'conced': 0.021076969986708317,
'violenc': 0.035669401563315412,
'fuell': 0.013197981377186868,
'antikurdish': 0.015157108319788976,
'feel': 0.031494546794920912,
'throughout': 0.031723341506223676,
'she': 0.013268580938863406,
'attempt': 0.052867977595169625,
'lynch': 0.0059149073505722072,
'season': 0.052629624051708346,
'western': 0.044878102278052674,
'sakarya': 0.0026218308932960573,
'pkks': 0.0063847639880995925,
'eas': 0.0063367804321689449,
'restrict': 0.012734764116991273,
'languag': 0.010711781399490085,
'offer': 0.014020416293548768,
'amnesti': 0.0072036689762947342,
'allow': 0.061589852722888118,

'their': 0.022631372791681051,
'seek': 0.0080759457747161378,
'asylum': 0.01728373206383664,
'solv': 0.013862286341976357,
'anyon': 0.060528191724240489,
'embolden': 0.032854836506676366,
'union': 0.075435720770520603,
'decis': 0.0031888977548963178,
'open': 0.0028622784401595128,
'explor': 0.0035606109318236682,
'possibl': 0.0070855865446990849,
'deal': 0.059809480062857502,
'accommod': 0.0064863121158863337,
'pressur': 0.027725712737235873,
'admit': 0.082553230342514258,
'made': 0.014558069556174297,
'mistak': 0.026339171153311702,
'cement': 0.0052505670716997792,
'justic': 0.029250724010866938,
'develop': 0.016185652714512291,
'ak': 0.003117869782356621,
'islam': 0.014128002768706798,
'credenti': 0.014178076148421972,
'play': 0.068656653679497692,
'well': 0.0096252478496403858,
'million': 0.083643588922186915,
'pious': 0.017296543084010733,
'dtps': 0.024268822900578434,
'rival': 0.033693131958248843,
'peopl': 0.061525176276120058,
'panacea': 0.014230070326000717,
'separat': 0.0009372146935607465,
'cemil': 0.011276729368395581,
'oter': 0.034537689935995528,
'tribal': 0.0075893871994449288,
'lost': 0.0019842334960452185,
'men': 0.011432933004031639,
'hope': 0.011717849665863894,
'fade': 0.014793104166983248,
'bigwig': 0.015820622629179361,
'franc': 0.035988626180110922,
'presid': 0.045277931800428044,
'nicola': 0.028515694570342458,
'sarkozi': 0.01313970478258422,
'lobbi': 0.0049854581211303131,
'entri': 0.026836189054058473,
'euinspir': 0.0091845193919930547,
'win': 0.015936969480356521,

'prize': 0.012575683231195586,
'date': 0.0074737019777277236,
'quiet': 0.015843116959864935,
'roll': 0.0069771742439886228,
'felt': 0.03617378543818682,
'local': 0.0069088681845389556,
'mayor': 0.030016812029233553,
'entir': 0.082068964710281425,
'council': 0.049475871699102052,
'bar': 0.0067555781602801467,
'offic': 0.027013793016318476,
'communic': 0.013815133846521666,
'constitu': 0.0073811452573763359,
'abdullah': 0.004119598045528217,
'demirba': 0.007340215360129302,
'annoy': 0.013864263693642031,
'author': 0.0060778395487736479,
'erect': 0.0086370460309564186,
'monument': 0.030407855431409822,
'memori': 0.020994572131494121,
'ahmet': 0.0052803811116197766,
'kaymaz': 0.0071651933284559745,
'lorri': 0.01779024068144755,
'driver': 0.043788577376833351,
'12yearold': 0.0040650695128799305,
'son': 0.021847317674531193,
'outsid': 0.017529576697194828,
'home': 0.0016375906564379094,
'town': 0.010606715775405789,
'kiziltep': 0.027981413914891208,
'ground': 0.016742325143472146,
'terrorist': 0.0052758363277222962,
'four': 0.029451058366121243,
'special': 0.031925258990630812,
'implic': 0.025276455495873189,
'kill': 0.024095659776872003,
'exoner': 0.00031280175659535161,
'court': 0.10646874473193775,
'april': 0.0014035580446183945,
'rare': 0.028385071397338674,
'tahir': 0.014425347089047571,
'elci': 0.017723108339297744,
'defend': 0.052484676057768943,
'famili': 0.085505970313741575,
'prison': 0.042400722147987735,
'criticis': 0.0038262729675150112,
'meanwhil': 0.010470152369802883,
'tortur': 0.015553786309488954,

'risen': 0.015181228141969269,
'regul': 0.0086891100432368792,
'detaine': 0.01574726085610131,
'deni': 0.02827668656175478,
'access': 0.05717698391801284,
'dure': 0.099423592388990956,
'24': 0.014033028642380714,
'hour': 0.036909200603161899,
'interrog': 0.015828359667124803,
'polic': 0.0012530856785760859,
'thing': 0.0097356229193228776,
'swell': 0.011169741682602079,
'rank': 0.023630253040239493,
'troubl': 0.019727182856360017,
'rebel': 0.0090744462189222656,
'tacticsset': 0.013065445060616033,
'landmin': 0.020203284482057094,
'plant': 0.023366420084699357,
'explosiveshav': 0.0045386418514094156,
'caught': 0.025062850717957923,
'guard': 0.0090489873200364727,
'gul': 9.8177073491697935e-05,
'foreign': 0.11526450918582162,
'reveal': 0.024285425341437506,
'tonn': 0.020783322572882974,
'plastic': 0.018737762156268302,
'explos': 0.015716945469188687,
'smuggl': 0.01823509569588299,
'alon': 0.030265842496571292,
'compromis': 0.029499844582829962,
'democraci': 0.066872325282089626,
'name': 0.005782446456750137,
'terror': 0.044560332037172634,
'trap': 0.0092000995267516334,
'each': 0.013202997871757199,
'turk': 0.003887787532996784,
'enthusiasm': 0.018735332011117451,
'preserv': 0.0082287454159342341,
'un': 0.0011741490177676068,
'departur': 0.021254097375717998,
'ingushetia': 0.012326438499694928,
'even': 0.033096832091617311,
'standard': 0.0053907532388817892,
'caucasus': 0.013146031614239266,
'littlenot': 0.014242079041012809,
'rocketpropel': 0.0020130707074445153,
'grenad': 0.0018048280033262222,
'march': 0.030065150106864433,

'nation': 0.015737240206846678,
'compound': 0.0057644397452198915,
'nazran': 0.0010407989801819008,
'citi': 0.034615091128256247,
'republ': 0.06201414613280265,
'russia': 0.044670912445833072,
'poorest': 0.0026525474787022201,
'host': 0.0076761323317125311,
'thousand': 0.04734814412358359,
'refuge': 0.0089008769942453807,
'safe': 0.02068170647928352,
'base': 0.020950276645974271,
'intern': 0.083906764335963863,
'work': 0.029369595745399869,
'neighbour': 0.031544444494095381,
'chechnya': 0.019874930631551158,
'bombard': 0.0057134371192345866,
'gave': 0.048411049286936139,
'lie': 0.021473711961888735,
'kremlin': 0.028307245044151042,
'claim': 0.0082419923505431796,
'life': 0.046914694872171453,
'normal': 0.010147312698690032,
'stori': 0.022753867879978275,
'organis': 0.094501972541513088,
'food': 0.04456914707235872,
'programm': 0.01600162971427559,
'unicef': 0.0052530947141599272,
'expat': 0.0036730504259839462,
'staff': 0.044412519939181286,
'vladikavkaz': 0.019469277458485829,
'aspect': 0.0091909842858328029,
'economist': 0.015584338506626074,
'understand': 0.031694298133080015,
'follow': 0.030786778081803069,
'disput': 0.020908703652419602,
'money': 0.052516335676752728,
'depart': 0.024050712750169519,
'bit': 0.061183962750119517,
'ingush': 0.01022209555955932,
'interior': 0.020884183956333038,
'respons': 0.0066649830798950743,
'appar': 0.03445327151726324,
'outfit': 0.012918242125588826,
'side': 0.035859019252710116,
'disagr': 0.00014124529845463852,
'poohpooh': 0.0038444693323305703,
'link': 0.022002456853086275,

'halfars': 0.022422428216960437,
'unnik': 0.011941332161344883,
'heard': 0.04349628112565565,
'protector': 0.00286818896878242,
'harass': 0.0063070526361774194,
'them': 0.022615882908834973,
'curious': 0.0067991228469713163,
'narrowli': 0.0087553343990016269,
'miss': 0.016799780814237168,
'build': 0.014019354135410814,
'team': 0.13122883354611961,
'unlik': 0.035076753291855493,
'cop': 0.0043417149161252867,
'sorri': 0.015416040334549982,
'tale': 0.047417554253006082,
'infam': 0.00047115221493665302,
'feder': 0.017841213841113174,
'servic': 0.040865419577013409,
'serious': 0.011545099792697756,
'kidnap': 0.013153369988004182,
'avoid': 0.053561258227901656,
'travel': 0.0074646475151446358,
'exempl': 0.05890392145154269,
'en': 0.0057981495113202412,
'rout': 0.016393819171771087,
'sympathi': 0.011826771587942991,
'littl': 0.031264455606985986,
'els': 0.024623378420552008,
'wonder': 0.0033736480362922651,
'bomb': 0.025696240542456592,
'commonplac': 0.014430691391035705,
'driven': 0.0034896367041111521,
'dizzi': 0.021841385061180971,
'mix': 0.014537829309211764,
'polit': 0.14894811062304103,
'religion': 0.01853830873793625,
'reveng': 0.016110147579352741,
'snatch': 0.021320027860815924,
'perpetr': 0.015096819950725728,
'chao': 0.01115204132760964,
'oilfuel': 0.0074263481801067023,
'bravado': 0.0056254508874538124,
'often': 0.040185357740531578,
'meet': 0.030647048624893859,
'reli': 0.0038341780748414401,
'instead': 0.012896619301715932,
'chariti': 0.0061509779749945341,
'routin': 0.023249587645278239,

'accus': 0.07147647274801322,
'latter': 0.019855079758740752,
'espionag': 0.012371440938907362,
'russiaa': 0.011459971842269441,
'perman': 0.013226472115110952,
'councilmay': 0.0047852150851273043,
'unabl': 0.034528646150255624,
'protect': 0.0048320813131957577,
'servant': 0.0068401018089320397,
'unwil': 0.0016413391020433943,
'hard': 0.074797549047296408,
'corrupt': 0.019077900169556065,
'wish': 0.036692598491070365,
'bless': 0.015082007482797912,
'small': 0.01771745035379382,
'fear': 0.070188794998675946,
'swamp': 0.0049991988127402994,
'migrant': 0.003140261644543087,
'event': 0.04387554053836009,
'await': 0.026667261029945304,
'malta': 0.010444503158607671,
'eager': 0.00068318154592501422,
'start': 0.059248404707196604,
'oper': 0.072403784714877434,
'nautilus': 9.3519139501350382e-05,
'ii': 0.0031296952253085202,
'pilot': 0.0055103625278114486,
'scheme': 0.043671363028119595,
'patrol': 0.001591151841467365,
'boat': 0.00042125658677355331,
'aircraft': 0.013281561694484696,
'converg': 0.020633698467970689,
'sea': 0.017739321163392843,
'libya': 0.02184721304509632,
'keep': 0.032120997112699862,
'bay': 0.010047857821149177,
'tide': 0.014232074733636267,
'eus': 0.0002628573917986826,
'smallest': 0.01043977272054819,
'most': 0.023334711289113231,
'crowd': 0.0011343690691701519,
'horn': 0.0026459710731741798,
'africa': 0.048932191121910655,
'overload': 0.00041259713929659851,
'unseaworthi': 0.012423173209434835,
'vessel': 0.0080977735535440343,
'lot': 0.037360641803616389,
'die': 0.029800822560156141,

'reach': 0.010561733742461589,
'safeti': 0.0032469876547537259,
'bodi': 0.017189266865297221,
'recov': 0.0099502022639726233,
'peril': 0.0022399005987451511,
'attitud': 0.0028036503263661285,
'maltes': 0.0048725231364357101,
'laid': 0.0056642101871965302,
'bare': 0.015952825385204834,
'27': 0.0015183263682211898,
'cling': 0.0054845060135316698,
'tuna': 0.01551033077308886,
'pen': 0.028638318147278852,
'attach': 0.0089865425402035886,
'malteseown': 0.00013554450021034277,
'water': 0.0059693654629262736,
'owner': 0.014211362762241804,
'might': 0.060562834844998643,
'solut': 0.0058431101583434663,
'suppli': 0.042461588319936863,
'withhold': 0.017426731743754538,
'cooper': 0.026868311148097256,
'giuliano': 0.0074570420240568673,
'amato': 0.0049601502792367847,
'exercis': 0.0014666795143690159,
'pointless': 0.00046091935861759532,
'unless': 0.00667547191309788,
'involv': 0.030334558120175274,
'repli': 0.0071425646325650781,
'invit': 0.0046541343494586338,
'frontex': 0.0013353866696549494,
'charg': 0.08094597902444374,
'cannot': 0.050044581950949497,
'take': 0.030070765093417044,
'part': 0.069669341920963454,
'send': 0.02283182076278276,
'spokesman': 0.0085729362219436143,
'inform': 0.055585763084783656,
'what': 0.035028587019413457,
'illeg': 0.011219356522382436,
'lifejacket': 0.0089585146769442861,
'distress': 0.0069884076846195714,
'rescu': 0.03374499951974351,
'taken': 0.029379362068597255,
'nearest': 0.0069273543257807393,
'placeperhap': 0.0093292843602097338,
'lampedusa': 0.014129657396977734,
'pantelleria': 0.021705216734900026,

```

'island': 0.0054316926553493634,
'itself': 0.04943573549277,
'deter': 0.0041107693535566651,
'migrat': 0.002220286299763645,
'safer': 0.0071599152953893833,
'antiimmigr': 0.022776576089032999,
...}

```

```

In [25]: from wordcloud import WordCloud
import matplotlib.pyplot as plt

from wordcloud import (WordCloud, get_single_color_func)
import matplotlib.pyplot as plt

class SimpleGroupedColorFunc(object):
    """Create a color function object which assigns EXACT colors
    to certain words based on the color to words mapping

    Parameters
    -----
    color_to_words : dict(str -> list(str))
        A dictionary that maps a color to the list of words.

    default_color : str
        Color that will be assigned to a word that's not a member
        of any value from color_to_words.
    """

    def __init__(self, color_to_words, default_color):
        self.word_to_color = {word: color
                               for (color, words) in color_to_words.items()
                               for word in words}

        self.default_color = default_color

    def __call__(self, word, **kwargs):
        return self.word_to_color.get(word, self.default_color)

wordcloud = WordCloud(collocations=False).generate_from_frequencies(wordslist)

color_to_words = {
    # words below will be colored with a green single color function
    'green': vocablist[:,0][positives],
    # will be colored with a red single color function

```

```
'red': vocablist[:,0][negatives]
}

# Words that are not in any of the color_to_words values
# will be colored with a grey single color function
default_color = 'grey'

# Create a color function with single tone
grouped_color_func = SimpleGroupedColorFunc(color_to_words, default_color)

# Create a color function with multiple tones
grouped_color_func = GroupedColorFunc(color_to_words, default_color)

# Apply our color function
wordcloud.recolor(color_func=grouped_color_func)

# Plot
plt.figure(figsize = (15, 20))
plt.imshow(wordcloud)
plt.axis("off")
plt.show()
```



- 4 We see that some color words are more likely to come from one article vs. the other.**

```
In [101]: print(Xtrain.shape)
```

(580, 26048)

```
In [109]: C = [10, 1, .1, .001]
```

```
for c in C:
    clf = LogisticRegression(penalty='l1', C=c)
    clf.fit(Xtrain, Ytrain)
    print('C:', C)
    print('Coefficient of each feature:', clf.coef_)
    print('Training accuracy:', clf.score(Xtrain, Ytrain))
    print('Test accuracy:', clf.score(Xtest, Ytest))
    print('')
```

```
clf = LogisticRegression(penalty='l1', C=1)
clf.fit(Xtrain, Ytrain)
print('C:', C)
print('Coefficient of each feature:', clf.coef_)
print('Training accuracy:', clf.score(Xtrain, Ytrain))
print('Test accuracy:', clf.score(Xtest, Ytest))
print('')
```

```
clf = LogisticRegression(penalty='l1', C=.1)
clf.fit(Xtrain, Ytrain)
print('C:', C)
print('Coefficient of each feature:', clf.coef_)
print('Training accuracy:', clf.score(Xtrain, Ytrain))
print('Test accuracy:', clf.score(Xtest, Ytest))
print('')
```

```
clf = LogisticRegression(penalty='l1', C=.001)
clf.fit(Xtrain, Ytrain)
print('C:', C)
print('Coefficient of each feature:', clf.coef_)
print('Training accuracy:', clf.score(Xtrain, Ytrain))
print('Test accuracy:', clf.score(Xtest, Ytest))
print('')
```

```
/Users/ericacollins/anaconda3/lib/python3.6/site-packages/sklearn/utils/validation.py:578: Data
y = column_or_1d(y, warn=True)
```

```
C: [10, 1, 0.1, 0.001]
```

```
Coefficient of each feature: [[ 0.          0.         -0.00949548 ...,  0.          0.
```

```
Training accuracy: 1.0
```

```
Test accuracy: 0.951724137931
```

```
C: [10, 1, 0.1, 0.001]
Coefficient of each feature: [[ 0.  0.  0. ...,  0.  0.  0.]]
Training accuracy: 1.0
Test accuracy: 0.937931034483
```

```
C: [10, 1, 0.1, 0.001]
Coefficient of each feature: [[ 0.  0.  0. ...,  0.  0.  0.]]
Training accuracy: 0.95
Test accuracy: 0.944827586207
```

```
C: [10, 1, 0.1, 0.001]
Coefficient of each feature: [[ 0.  0.  0. ...,  0.  0.  0.]]
Training accuracy: 0.727586206897
Test accuracy: 0.675862068966
```

```
C: [10, 1, 0.1, 0.001]
Coefficient of each feature: [[ 0.  0.  0. ...,  0.  0.  0.]]
Training accuracy: 1.0
Test accuracy: 0.937931034483
```

```
C: [10, 1, 0.1, 0.001]
Coefficient of each feature: [[ 0.  0.  0. ...,  0.  0.  0.]]
Training accuracy: 0.95
Test accuracy: 0.944827586207
```

```
C: [10, 1, 0.1, 0.001]
Coefficient of each feature: [[ 0.  0.  0. ...,  0.  0.  0.]]
Training accuracy: 0.727586206897
Test accuracy: 0.675862068966
```

```
In [58]: import numpy as np
         from sklearn.model_selection import KFold # import KFold
         X = np.array([[1, 2], [3, 4], [1, 2], [3, 4]]) # create an array
         y = np.array([1, 2, 3, 4]) # Create another array
         kf = KFold(n_splits=2) # Define the split - into 2 folds
         kf.get_n_splits(X) # returns the number of splitting iterations in the cross-validator
         print(kf)
```

```
KFold(n_splits=2, random_state=None, shuffle=False)
```

```
In [57]: import pandas as pd
         from matplotlib import pyplot as plt
         for train_index, test_index in kf.split(X):
             print('TRAIN:', train_index, 'TEST:', test_index)
```

```
Xtrain, Xtest = X[train_index], X[test_index]
Ytrain, Ytest = y[train_index], y[test_index]
```

```
TRAIN: [2 3] TEST: [0 1]
TRAIN: [0 1] TEST: [2 3]
```

```
In [ ]:
```

```
In [73]: C = [10, 1, .1, .001]
```

```
for c in C:
    clf = LogisticRegression(penalty='l1', C=c)
    clf.fit(Xtrain, Ytrain)
    print('C:', C)
    print('Coefficient of each feature:', clf.coef_)
    print('Training accuracy:', clf.score(Xtrain, Ytrain))
    print('Test accuracy:', clf.score(Xtest, Ytest))
    print('')
```

```
C: [10, 1, 0.1, 0.001]
Coefficient of each feature: [[ 1.79114043  0.          ]]
Training accuracy: 1.0
Test accuracy: 0.0
```

```
C: [10, 1, 0.1, 0.001]
Coefficient of each feature: [[ 0.  0.]]
Training accuracy: 0.5
Test accuracy: 0.0
```

```
C: [10, 1, 0.1, 0.001]
Coefficient of each feature: [[ 0.  0.]]
Training accuracy: 0.5
Test accuracy: 0.0
```

```
C: [10, 1, 0.1, 0.001]
Coefficient of each feature: [[ 0.  0.]]
Training accuracy: 0.5
Test accuracy: 0.0
```

```
In [ ]: totalx = np.concatenate([Xtrain, Xtest])
        print(totalx.shape)
```

```
In [115]: totaly = np.concatenate([Ytrain, Ytest])
          print(totaly.shape)
          print('# Tuning hyper-parameters for %s')
```

```
(725, 1)
# Tuning hyper-parameters for %s
```

```
In [ ]:
```

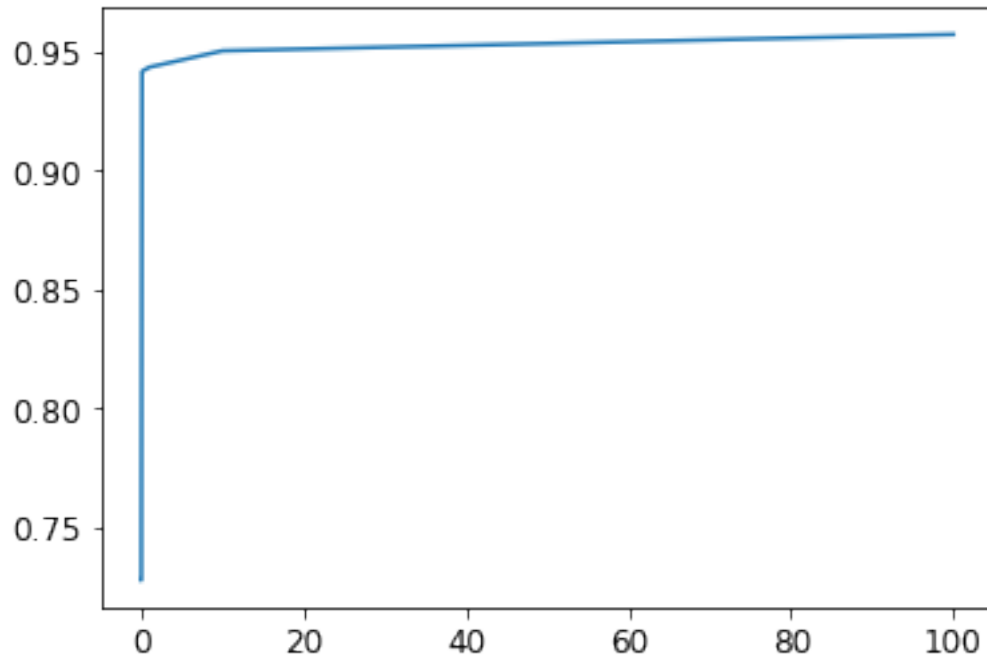
```
In [145]: from sklearn.model_selection import cross_val_score
          from statistics import mean
          result_array = np.array([])
          C = [100, 10, 1, .1, .01, .001, .0001]
          listx = []
          for c in C:
              clf = LogisticRegression(penalty='l1', C=c)
              scores = cross_val_score(clf, Xtrain, Ytrain.ravel(), cv=5)
              print(scores)
              result_array = np.append(result_array, mean(scores))
```

```
[ 0.95726496  0.94871795  0.94827586  0.96521739  0.96521739]
[ 0.94871795  0.93162393  0.96551724  0.94782609  0.95652174]
[ 0.94871795  0.90598291  0.97413793  0.94782609  0.93913043]
[ 0.93162393  0.93162393  0.96551724  0.94782609  0.93043478]
[ 0.72649573  0.72649573  0.72413793  0.73043478  0.73043478]
[ 0.72649573  0.72649573  0.72413793  0.73043478  0.73043478]
[ 0.72649573  0.72649573  0.72413793  0.73043478  0.73043478]
```

```
In [146]: print(result_array)
```

```
[ 0.95693871  0.95004139  0.94315906  0.94140519  0.72759979  0.72759979
 0.72759979]
```

```
In [147]: import matplotlib.pyplot as plt
          plt.plot(C, result_array)
          plt.show()
```

5 The error is better with higher values?

```
In [ ]: clf = LogisticRegression(penalty='l1', C=c).fit(Xtrain, Ytrain)
```