Matlab - 2

Mikhail Andreev

November 2, 2015

1 San Francisco Crime Prediction

1.1 Part a

Histogram of incidents per day of the week

Histogram of incidents per district
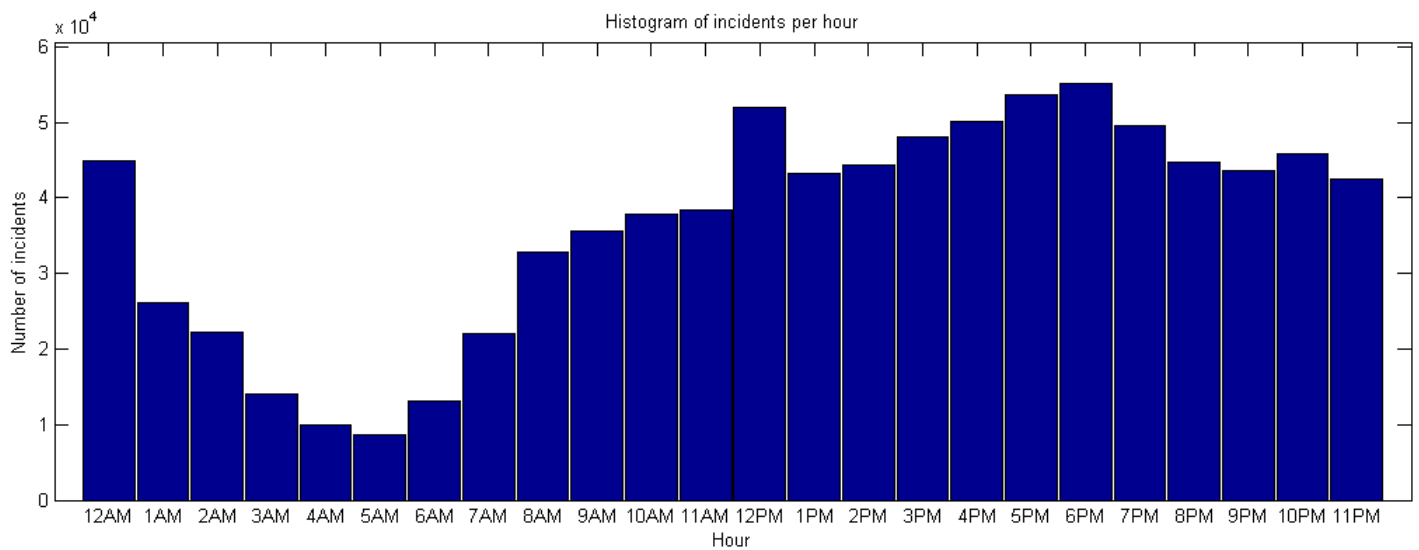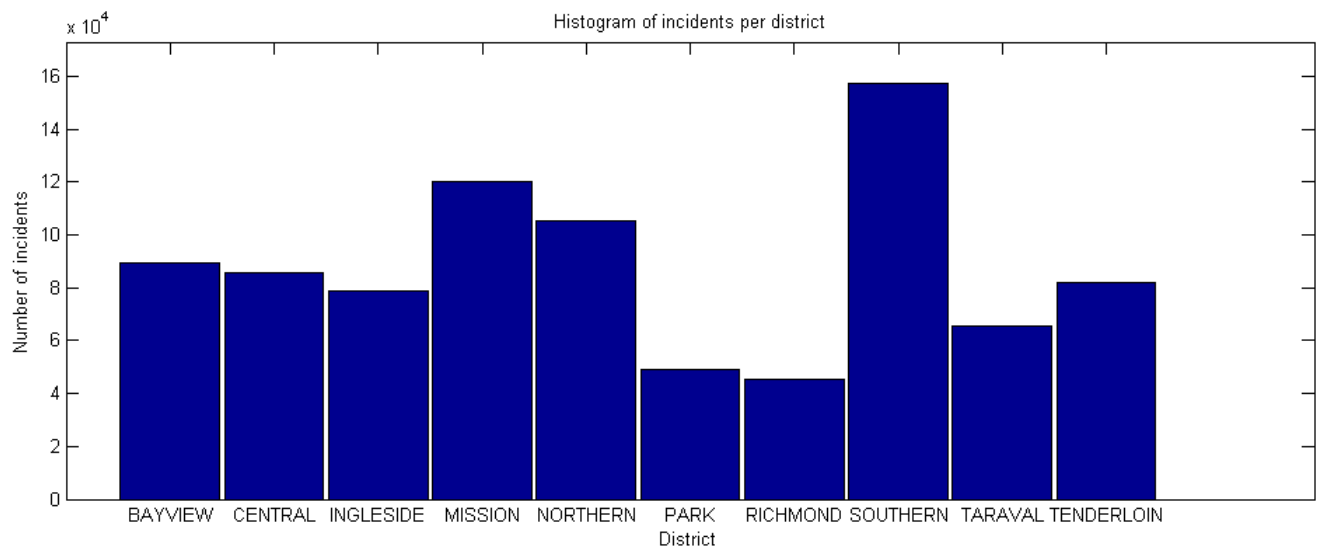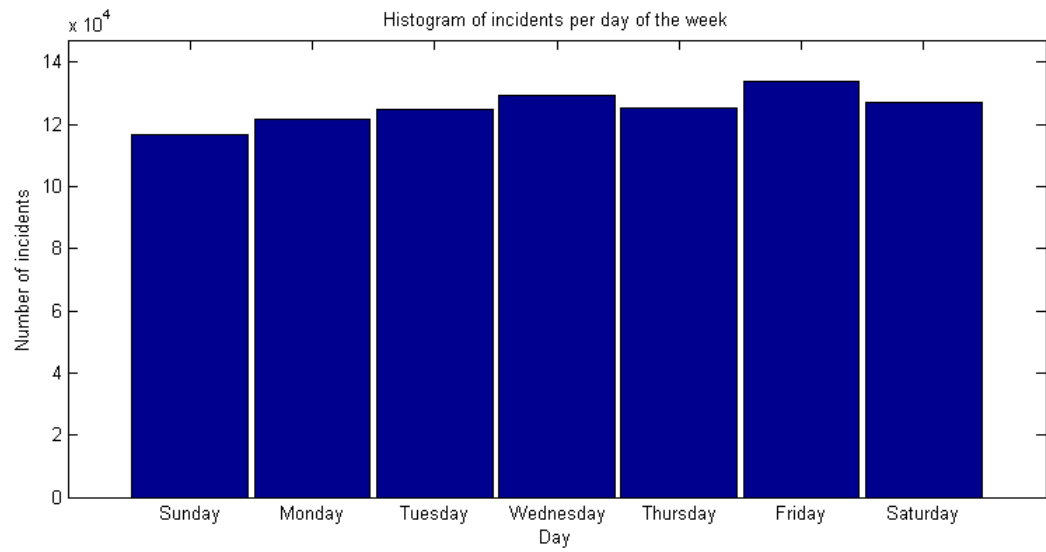
The list of crimes with the hour they are most likely to occur.

Crime	Hour
ARSON	0
ASSAULT	0
BAD_CHECKS	12
BRIBERY	17
BURGLARY	17
DISORDERLY_CONDUCT	6
DRIVING_UNDER_THE_INFLUENCE	0
DRUG_NARCOTIC	14
DRUNKENNESS	0
EMBEZZLEMENT	0
EXTORTION	0
FAMILY_OFFENSES	15
FORGERY_COUNTERFEITING	0
FRAUD	0

GAMBLING	13
KIDNAPPING	0
LARCENY_THEFT	18
LIQUOR_LAWS	17
LOITERING	17
MISSING_PERSON	8
NON_CRIMINAL	12
OTHER_OFFENSES	17
PORNOGRAPHY_OBSCENE_MAT	14
PROSTITUTION	22
RECOVERED_VEHICLE	12
ROBBERY	21
RUNAWAY	18
SECONDARY_CODES	12
SEX_OFFENSES_FORCIBLE	0
SEX_OFFENSES_NON_FORCIBLE	0
STOLEN_PROPERTY	16
SUICIDE	18
SUSPICIOUS_OCC	12
TREA	5
TRESPASS	6
VANDALISM	18
VEHICLE_THEFT	18
WARRANTS	17
WEAPON_LAWS	16

The crime most likely to occur in BAYVIEW is OTHER OFFENSES
The crime most likely to occur in CENTRAL is LARCENY and THEFT
The crime most likely to occur in INGLESIDE is OTHER OFFENSES
The crime most likely to occur in MISSION is OTHER OFFENSES
The crime most likely to occur in NORTHERN is LARCENY and THEFT
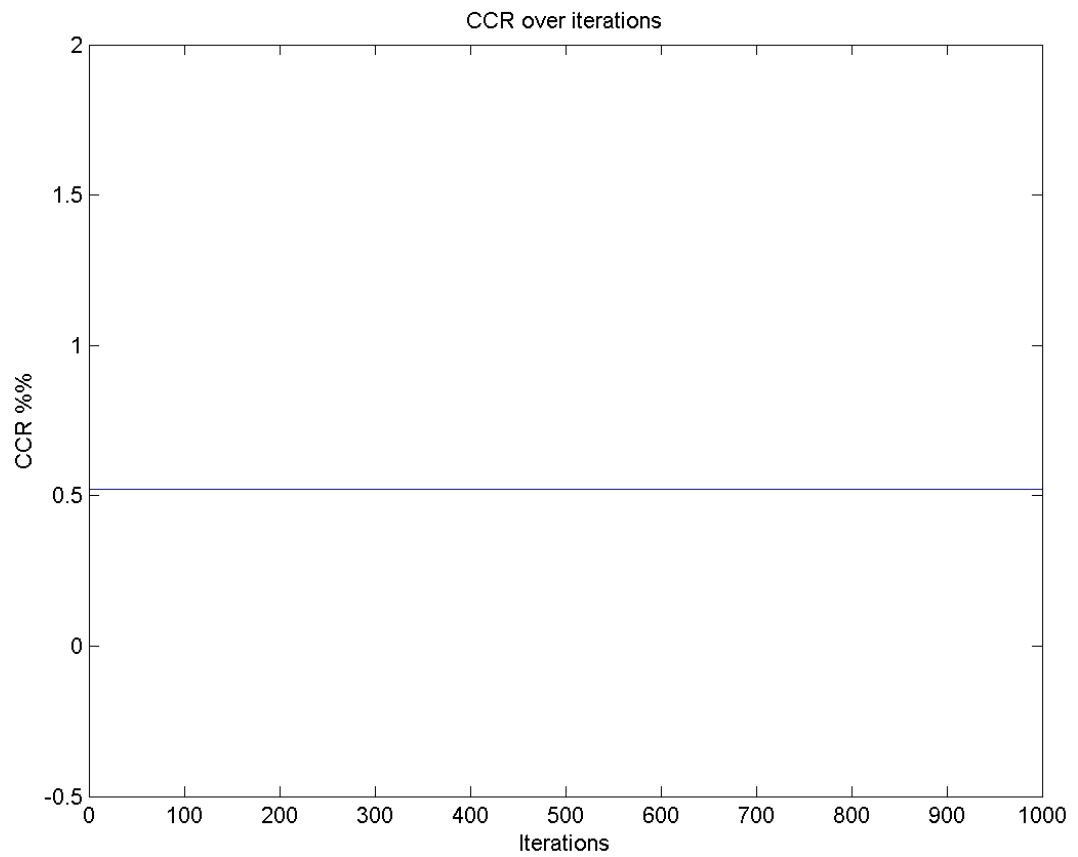The crime most likely to occur in PARK is LARCENY and THEFT
The crime most likely to occur in RICHMOND is LARCENY and THEFT
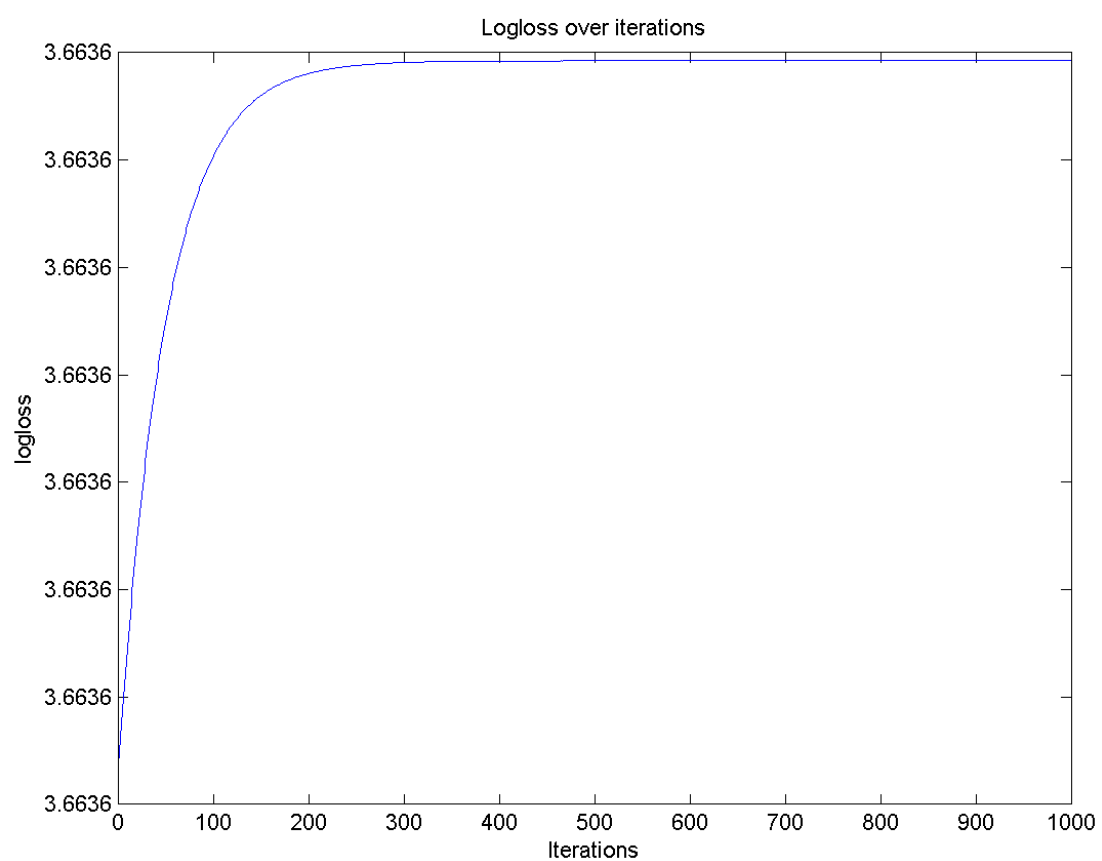The crime most likely to occur in SOUTHERN is LARCENY and THEFT
The crime most likely to occur in TARAVAL is LARCENY and THEFT
The crime most likely to occur in TENDERLOIN is DRUG and NARCOTIC
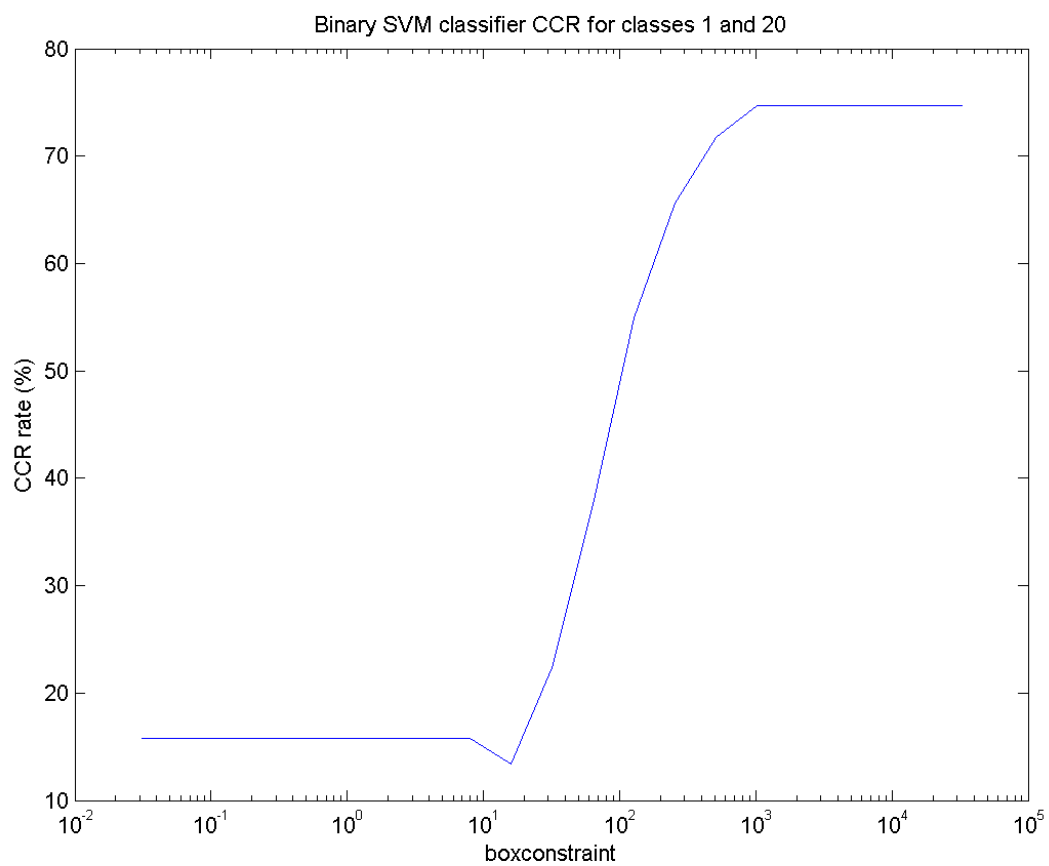
1.2 PART B

This graph indicates that the CCR rate did not noticeably change over the different iterations of the parameters. Unfortunately, this also indicates the CCR rate was very low for the experiment. The end result was only .5% correct detection.

Logloss over iterations

1.3 Part c

2 SVM Classifier for Text Documents

2.1 Part a

Binary SVM classifier CCR for classes 1 and 20

$C^* = 2^{10}$

$CCR = 80.32\%$

2.2 Part b

Binary SVM classifier CCR for classes 1 and 20 using RBF
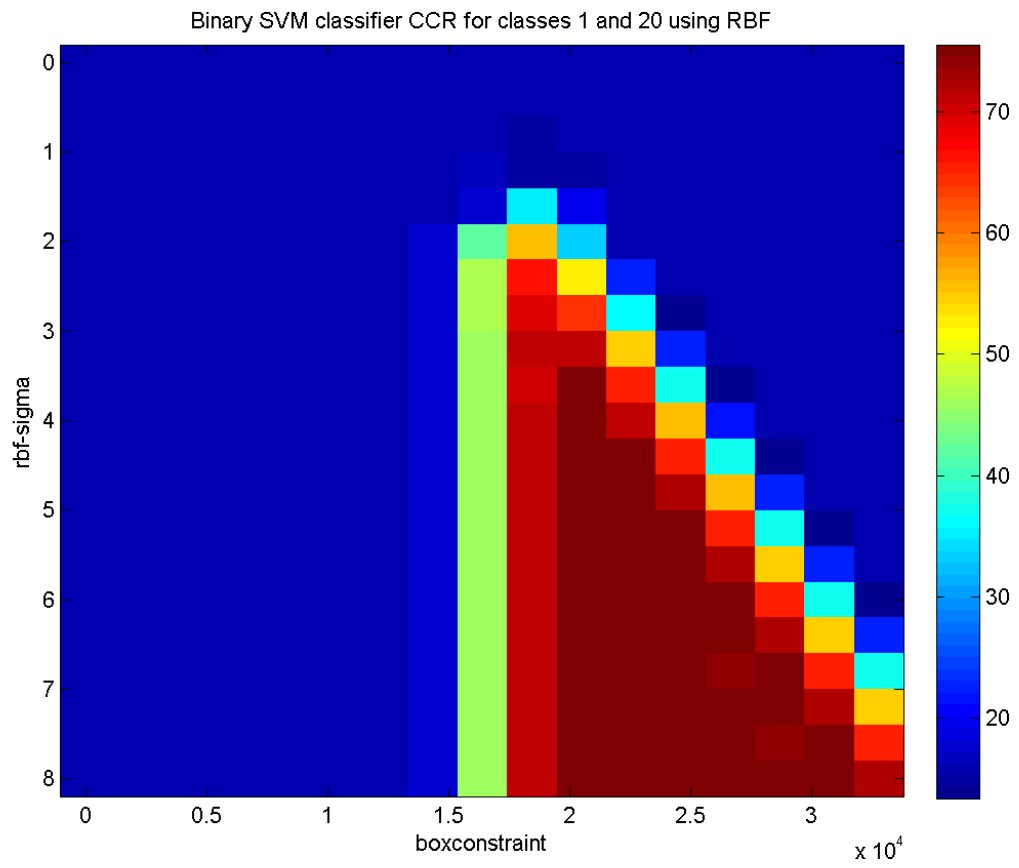
$C^* = 2^{11}$

$rbf - sigma - 2^5$

$CCR = 79.44\%$

Binary SVM classifier CCR for classes 17 and the rest
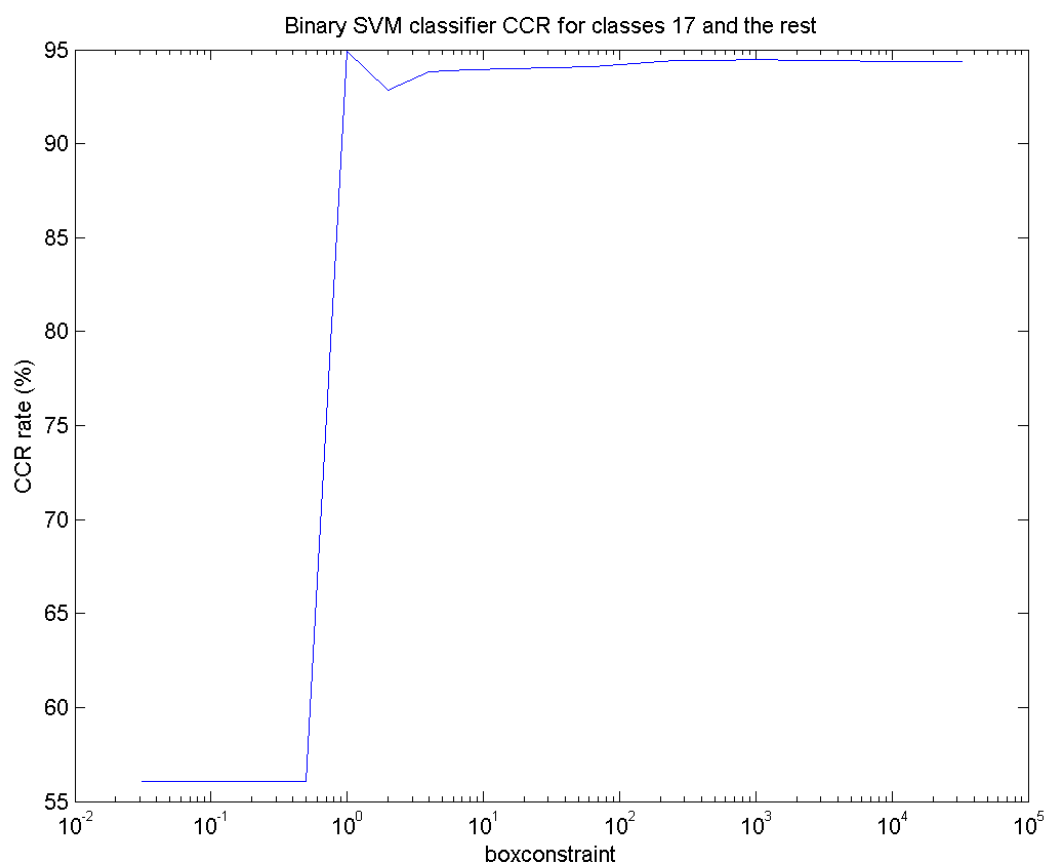
$C^* = 2^0$

$CCR = 95.78\%$

Confusion Matrix:

	True Condition	
Predicted	217	147
Condition	170	6971

2.4 Part d

Binary SVM classfier Precision, Recall, and F-Score
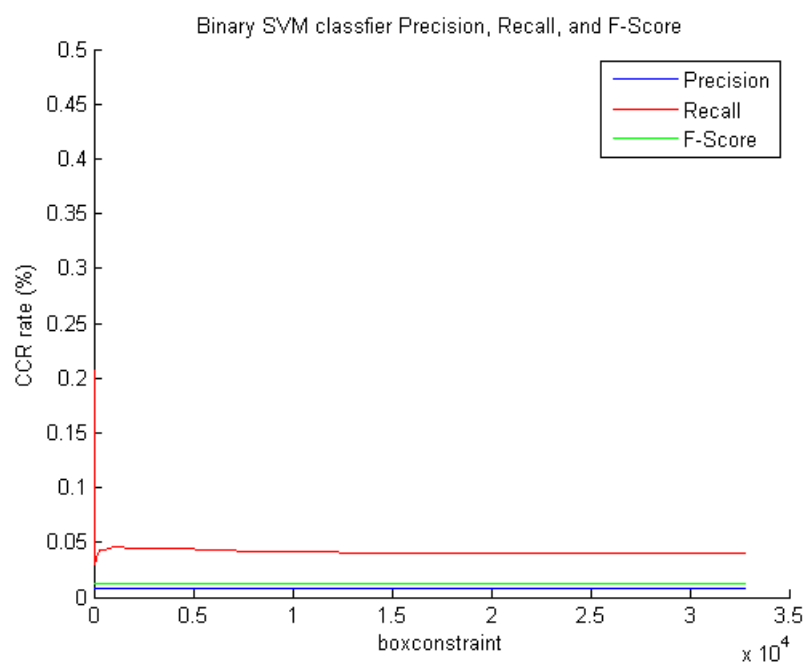
Best value of C as determined by the precision and the F-score is 2^0

Confusion Matrix:

	True Condition	
Predicted	902.6	1009.4
Condition	908.8	2137.8

2.5 Part e

2.6 Part f