

Question 1

DS1.csv attached along with Jupyter Notebook files.

Question 2

Best fit Accuracy: 0.945

Precision: 0.9556

Recall: 0.933

F-Measure: 0.944

Coefficients Learnt:

U1:

[1.92486305 1.88574537 1.91903303 1.9023876 1.87220534 1.92675551 1.90384851
1.93631247 1.92383374 1.88307188 1.93366121 1.8939727 1.8565985 1.92555162
1.90964703 1.86961943 1.91901221 1.90444018 1.92062611 1.86677443]

U2:

[1.2578635 1.33369734 1.22781812 1.24713771 1.2745293 1.2742541 1.29876692
1.24321133 1.23592936 1.29741145 1.26109714 1.26499492 1.27306806 1.24368497
1.26658048 1.24846667 1.20458703 1.26859286 1.28606356 1.29473774]

Pi:

0.5

S:

[[8.0615456 5.54023354 6.33323801 5.24674662 6.02534845 6.29813792 4.80638065 5.56427935 5.08785888 5.306586
4.06744829 5.34554859 7.19348875 6.16442614 6.23079925 6.14010753 5.9777847 5.79110427 5.75451037 6.05636299]

[5.54023354 6.82704627 5.45586708 4.35102419 5.54560142 5.62241695 4.38019284 3.99700092 4.17439172 5.0667105
3.42920118 4.63308771 5.91255215 5.22008792 5.51742136 5.27390571 5.68707093 5.15873333 5.42499206 5.36940271]

[6.33323801 5.45586708 7.46264998 4.91798934 5.95914007 6.76916316 4.71897136 4.91492352 4.97997905 5.21088771
3.37949685 4.76855865 6.49880727 5.31054049 6.24094858 6.08816915 6.36884001 5.08964861 4.86479941 5.21529275]

[5.24674662 4.35102419 4.91798934 5.8091975 5.35139632 4.55298434 3.75199064 4.44548935 3.42112162 4.32414227
2.84503784 4.22380196 5.97935023 4.89707982 4.84113141 5.13355756 4.70817118 4.64109447 4.00727385 5.89717266]

[6.02534845 5.54560142 5.95914007 5.35139632 7.24158198 5.56146878 5.17666702 4.52160611 4.82956956 5.29823818
4.2698059 5.14519563 6.3853618 6.09870671 6.16911122 6.39678468 6.09182489 5.31232918 5.61497257 5.93510321]

[6.29813792 5.62241695 6.76916316 4.55298434 5.56146878 6.90224511 4.56974591 5.00880734 4.84407908 5.47027041
3.09009518 4.91080693 6.51065036 5.13528913 5.94013165 6.12704514 6.04957686 5.10459406 4.92643946 5.32832003]

[4.80638065 4.38019284 4.71897136 3.75199064 5.17666702 4.56974591 5.24968557 4.0064217 4.07518813 4.47199267
3.1414642 4.43405465 4.84002363 4.30587025 4.82649804 5.41616197 4.45661031 3.8042805 4.78412201 4.01869939]

[5.56427935 3.99700092 4.91492352 4.44548935 4.52160611 5.00880734 4.0064217 6.1373358 3.65479466 5.08793196
2.66883256 4.97390283 6.33141276 5.31169997 4.98117821 6.29002344 4.6004594 5.04278742 4.97100901 4.61085804]

[5.08785888 4.17439172 4.97997905 3.42112162 4.82956956 4.84407908 4.07518813 3.65479466 5.01442827 4.69979339
3.21369765 4.14270938 5.13117672 4.75966807 5.02039635 5.25938176 5.1969007 4.11853029 4.36184046 4.15180785]

[5.306586 5.0667105 5.21088771 4.32414227 5.29823818 5.47027041 4.47199267 5.08793196 4.69979339 7.22993603
3.39593981 4.84423839 6.98703688 5.33810472 5.13071138 6.12028485 4.883369 5.39420652 4.71409546 5.47381829]

[4.06744829 3.42920118 3.37949685 2.84503784 4.2698059 3.09009518 3.1414642 2.66883256 3.21369765 3.39593981
3.4797634 3.02109906 3.81563439 4.11780763 3.75838935 3.83249693 3.75278219 3.32134872 3.92092593 3.58950132]

[5.34554859 4.63308771 4.76855865 4.22380196 5.14519563 4.91080693 4.43405465 4.97390283 4.14270938 4.84423839
3.02109906 5.34107908 6.1199687 5.14996843 4.905715 5.68254328 5.12847848 4.97629853 5.17278444 4.60451951]

[7.19348875 5.91255215 6.49880727 5.97935023 6.3853618 6.51065036 4.84002363 6.33141276 5.13117672 6.98703688
3.81563439 6.1199687 9.233844 6.42666165 6.36163125 6.70248276 6.08119386 6.44945735 6.03404879 7.43293555]

[6.16442614 5.22008792 5.31054049 4.89707982 6.09870671 5.13528913 4.30587025 5.31169997 4.75966807 5.33810472
4.11780763 5.14996843 6.42666165 7.19391908 6.07129249 6.41289369 5.85536464 5.73644682 6.16412539 5.70236574]

[6.23079925 5.51742136 6.24094858 4.84113141 6.16911122 5.94013165 4.82649804 4.98117821 5.02039635 5.13071138
3.75838935 4.905715 6.36163125 6.07129249 7.1183353 7.29991019 5.96146569 5.23855267 5.67257622 5.90237173]

[6.14010753 5.27390571 6.08816915 5.13355756 6.39678468 6.12704514 5.41616197 6.29002344 5.25938176 6.12028485
3.83249693 5.68254328 6.70248276 6.41289369 7.29991019 9.50012532 5.99164346 6.01459484 5.86142663 6.03040186]

[5.9777847 5.68707093 6.36884001 4.70817118 6.09182489 6.04957686 4.45661031 4.6004594 5.1969007 4.883369
3.75278219 5.12847848 6.08119386 5.85536464 5.96146569 5.99164346 6.84247695 5.28922571 5.47552669 5.04327693]

[5.79110427 5.15873333 5.08964861 4.64109447 5.31232918 5.10459406 3.8042805 5.04278742 4.11853029 5.39420652
3.32134872 4.97629853 6.44945735 5.73644682 5.23855267 6.01459484 5.28922571 6.03446742 4.99494465 5.23636401]

[5.75451037 5.42499206 4.86479941 4.00727385 5.61497257 4.92643946 4.78412201 4.97100901 4.36184046 4.71409546
3.92092593 5.17278444 6.03404879 6.16412539 5.67257622 5.86142663 5.47552669 4.99494465 6.61169287 5.08615366]

[6.05636299 5.36940271 5.21529275 5.89717266 5.93510321 5.32832003 4.01869939 4.61085804 4.15180785 5.47381829
3.58950132 4.60451951 7.43293555 5.70236574 5.90237173 6.03040186 5.04327693 5.23636401 5.08615366 8.1605727]]

w

[-14.88105119 8.69068246 5.56391428 3.26425236 9.81126017
4.62701308 -17.20608142 24.84481984 29.87743031 -9.49635768
13.62003589 12.47560219 -15.94523547 -13.28126465 5.92634042
-13.48856649 -29.98969146 7.04258343 0.43809373 5.12597768]

w0

-28.0790058468

Question 3

K = 1

Precision 0.55

Recall 0.568333333333

F-Measure 0.559016393443

K = 3

Precision 0.567260940032

Recall 0.583333333333

F-Measure 0.575184880855

K = 5

Precision 0.56765163297

Recall 0.608333333333

F-Measure 0.587288817377

K = 10

Precision 0.563897763578

Recall 0.588333333333

F-Measure 0.575856443719

K = 20

Precision 0.560747663551

Recall 0.6

F-Measure 0.579710144928

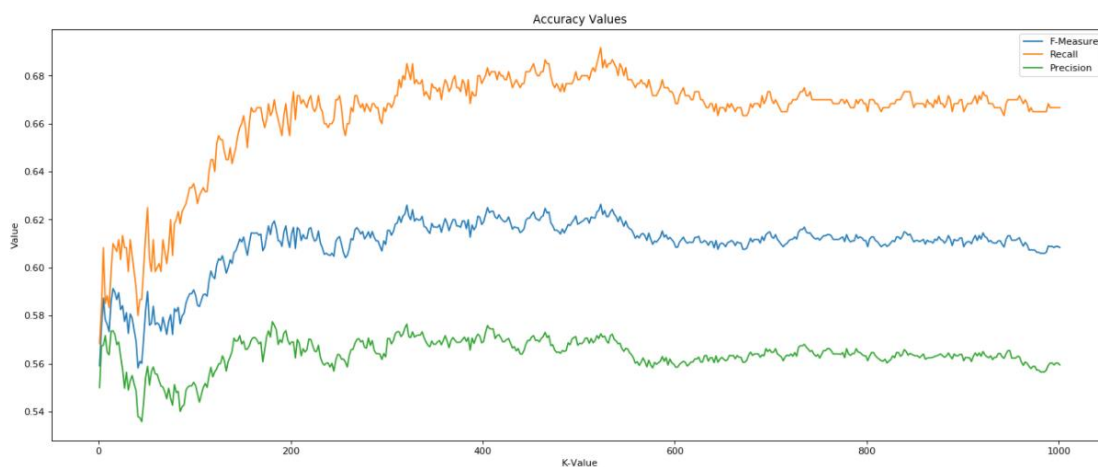
K = 50

Precision 0.547085201794

Recall 0.61

F-Measure 0.5768321513

Dataset 1 (1000 K-Values)



Highest Precision 0.577424023155 at K-value 181.0

Highest Recall 0.691666666667 at K-value 523.0

Highest F-Measure 0.62641509434 at K-value 523.0

The K-NN classifier performs significantly worse than the LDA classifier. In LDA, for dataset 1, the F-Measure was 0.944, while with K-NN, the best F-Measure is 0.626. This occurs at a K-value of 523.

K = 523

Precision 0.572413793103

Recall 0.691666666667

F-Measure 0.62641509434

As the graph above shows, the classifier gets better as we increase the K value but plateaus very quickly.

Question 4

DS2.csv attached along with Jupyter Notebook files

Question 5

LDA

Best fit Accuracy: 0.5008

Precision: 0.518

Recall: 0.462

F-Measure: 0.488

Coefficients Learnt:

U1

[0.93764887 0.95352548 0.98280782 0.96581479 0.9405326 0.98589368 0.99037942
0.94295805 1.01208443 0.95809837 0.99408319 0.94203745 0.99270284 0.93522284
0.97377322 0.96623913 0.99524936 0.88053059 0.96744888 1.00763805]

U2

[1.27801399 1.26026459 1.25649807 1.1953158 1.18132912 1.25296546 1.27737524
1.23315997 1.29595725 1.24760954 1.28463662 1.23443485 1.25919227 1.30971448
1.20792882 1.25958999 1.23267725 1.18948249 1.2307232 1.20313014]

Pi

0.4932142857142857

S

[[7.74324131 5.53826411 4.80422318 5.19106882 4.38149587 5.77851032 5.96042789 5.77341115 4.80976661 5.61645192
5.56964799 5.12133164 5.11325932 6.22405301 5.54769895 5.74641815 5.5907825 5.47768377 5.82994051 6.05508882]

[5.53826411 7.40642097 5.15966791 5.42794621 5.15655824 6.23381724 6.6674432 5.59808488 4.8620397 5.39814389
4.99027058 5.26934222 5.14625018 6.38510965 5.9482866 6.12075168 5.52195867 5.21411255 6.4572594 5.97094789]

[4.80422318 5.15966791 7.00668132 5.46398196 4.98015458 5.0631269 6.15181198 4.79675451 4.79650027 5.07554591
4.91756673 5.11812309 5.2605094 6.18995532 5.66755997 5.08480747 5.5946868 4.59423058 6.06041395 6.11030455]

[5.19106882 5.42794621 5.46398196 6.77555481 4.35517529 5.81017357 6.42240799 5.54753382 5.05591319 5.71599789
5.08923481 4.94816212 5.22234606 5.70975314 5.62340904 5.52393769 5.46723109 5.28241687 6.06665502 6.01003953]

[4.38149587 5.15655824 4.98015458 4.35517529 5.72739001 4.81054496 5.19655941 4.6081973 3.86223893 4.23655429
3.9048938 4.30007782 4.36545698 4.96806232 5.38550218 5.00925711 4.4679011 4.61709509 5.49633308 5.0819924]

[5.77851032 6.23381724 5.0631269 5.81017357 4.81054496 8.0684849 6.88325499 5.45315776 5.73134901 5.85334797
4.91488659 5.13825186 5.60970642 6.71042891 5.96008173 6.10372305 5.87174063 5.45712003 6.85033492 6.36401806]

[5.96042789 6.6674432 6.15181198 6.42240799 5.19655941 6.88325499 8.55370579 6.13010593 6.07773665 6.3252808
5.89462777 5.85682932 5.773275 7.11989578 6.2363136 6.42797801 6.10737386 5.70578963 7.38324013 6.5520106]

[5.77341115 5.59808488 4.79675451 5.54753382 4.6081973 5.45315776 6.13010593 6.69067566 4.86876409 5.38815381
4.7191107 4.87964828 4.47709229 5.66184842 5.50320526 5.45663839 5.12325971 5.74918118 5.80944472 5.67572518]

[4.80976661 4.8620397 4.79650027 5.05591319 3.86223893 5.73134901 6.07773665 4.86876409 6.17282152 4.86515266
4.46127341 4.52677367 4.5351026 5.3464588 4.9701882 4.67194974 4.97234959 4.72733966 5.13333709 5.10345778]

[5.61645192 5.39814389 5.07554591 5.71599789 4.23655429 5.85334797 6.3252808 5.38815381 4.86515266 7.18830075
4.9068553 4.84310299 5.14668468 6.02662757 5.50292109 5.37926306 5.34003365 4.76068485 6.5670097 5.67535619]

[5.56964799 4.99027058 4.91756673 5.08923481 3.9048938 4.91488659 5.89462777 4.7191107 4.46127341 4.9068553
6.13832537 5.13051043 4.73179207 5.22295059 5.2136892 4.82229298 4.96921509 4.70084859 5.77222257 5.55009613]

[5.12133164 5.26934222 5.11812309 4.94816212 4.30007782 5.13825186 5.85682932 4.87964828 4.52677367 4.84310299
5.13051043 6.57582194 4.67147735 5.4880663 5.63222321 5.18663151 5.41830888 5.01498706 5.84034532 5.46893551]

[5.11325932 5.14625018 5.2605094 5.22234606 4.36545698 5.60970642 5.773275 4.47709229 4.5351026 5.14668468
4.73179207 4.67147735 6.39679664 5.63751948 5.73158223 5.49280462 5.41969603 4.71882591 6.02632058 5.60816228]

[6.22405301 6.38510965 6.18995532 5.70975314 4.96806232 6.71042891 7.11989578 5.66184842 5.3464588 6.02662757
5.22295059 5.4880663 5.63751948 8.38312201 6.21037141 5.96046854 6.42653934 4.73240326 6.64189145 6.53731961]

[5.54769895 5.9482866 5.66755997 5.62340904 5.38550218 5.96008173 6.2363136 5.50320526 4.9701882 5.50292109
5.2136892 5.63222321 5.73158223 6.21037141 7.55994569 5.7002299 5.77097049 5.50139063 6.30921938 6.49249595]

[5.74641815 6.12075168 5.08480747 5.52393769 5.00925711 6.10372305 6.42797801 5.45663839 4.67194974 5.37926306
4.82229298 5.18663151 5.49280462 5.96046854 5.7002299 6.88079914 5.52457503 5.63464588 6.47442128 5.97082665]

[5.5907825 5.52195867 5.5946868 5.46723109 4.4679011 5.87174063 6.10737386 5.12325971 4.97234959 5.34003365
4.96921509 5.41830888 5.41969603 6.42653934 5.77097049 5.52457503 7.02469545 4.80695537 6.03993716 6.41603869]

[5.47768377 5.21411255 4.59423058 5.28241687 4.61709509 5.45712003 5.70578963 5.74918118 4.72733966 4.76068485
4.70084859 5.01498706 4.71882591 4.73240326 5.50139063 5.63464588 4.80695537 6.8136515 5.73276865 5.56272301]

[5.82994051 6.4572594 6.06041395 6.06665502 5.49633308 6.85033492 7.38324013 5.80944472 5.13333709 6.5670097
5.77222257 5.84034532 6.02632058 6.64189145 6.30921938 6.47442128 6.03993716 5.73276865 8.69608106 6.74079146]

[6.05508882 5.97094789 6.11030455 6.01003953 5.0819924 6.36401806 6.5520106 5.67572518 5.10345778 5.67535619
5.55009613 5.46893551 5.60816228 6.53731961 6.49249595 5.97082665 6.41603869 5.56272301 6.74079146 8.15131583]]

W

[0.02464284 -0.0170207 0.00066467 0.03149945 -0.02743455 0.020
43042
0.06893759 0.00012959 -0.03661041 -0.03594859 -0.06209093 -0.0196
6858
-0.03294264 -0.09641688 0.06302365 -0.01202011 0.03468251 -0.0632
0098
0.03414195 0.04172576]

w0

0.0646602492869

K-NN

K = 1

Precision 0.541864139021

Recall 0.554119547658

F-Measure 0.547923322684

K = 3

Precision 0.529886914378

Recall 0.529886914378

F-Measure 0.529886914378

K = 5

Precision 0.535773710483

Recall 0.520193861066

F-Measure 0.527868852459

K = 10

Precision 0.530405405405

Recall 0.507269789984

F-Measure 0.51857968621

K = 20

Precision 0.541254125413

Recall 0.529886914378

F-Measure 0.535510204082

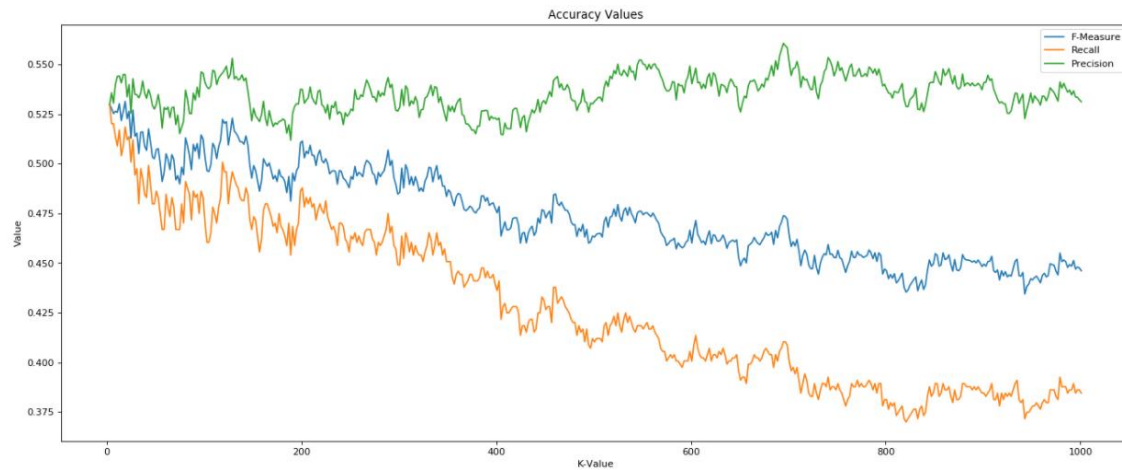
K = 50

Precision 0.533101045296

Recall 0.494345718901

F-Measure 0.512992455993

Dataset 2 (1000 K-Values)



K = 19

Precision 0.544991511036

Recall 0.518578352181

F-Measure 0.531456953642

The best performance for the second dataset under K-NN classifier occurs at K = 19. The F-Measure is 0.5314 at this value.

When we create a new dataset, one which has vectors that are random gaussian variables coming from different covariance matrices, the accuracy of our classifier decreases significantly. With the data coming from a single covariance matrix the F-Measure was 0.63, while in this case it is 0.53. This is a whole 10 percent off. That is an enormous number when it comes to accuracy of classifiers. The classifier has the option of choosing either 0 or 1. A coin flip would be able to classify the data with 50% accuracy. The F-measure would also be around this number. The classifier does barely better than a random coin flip. This is obviously not a good classifier.

Question 6

The LDA classifier performed superbly for the first dataset. It achieved more than 95% accuracy, and 94% recall. Meanwhile, for the second dataset, the accuracy was only 50%, and the F-measure was even lower—at 0.48. The accuracy of the LDA with the second dataset is just as accurate as a random coin flip. This is less than ideal. When the K-NN was tested on the first dataset, the F-Measure values dipped significantly. This is because the data is linearly separated in a 20-dimensional space, so our linear classifier can classify it with much greater accuracy. Because the two classes are near each other (they must be or else the results for the KNN would be better), the KNN does not do so well. It takes the nearest neighbors and votes to choose a class, but it seems as if the nearest neighbors aren't necessarily the ones from the same class. When the covariance matrices of the distributions can vary, the results get even worse. Because the data is no longer linearly distributed, it is no surprise that the LDA only has 50% accuracy, as

previously mentioned. The system essentially picks a line to split the data, and this boundary correctly classifies data as good as a random coin toss. The KNN does slightly better than a random coin toss, however, there is not that much improvement here either. The F-Measure only goes up to 0.54. The KNN of the second dataset is also worse than the first. This is presumably because the distribution of the second dataset can vary more, hence the distances to the nearest neighbors are even more widely distributed between classes.