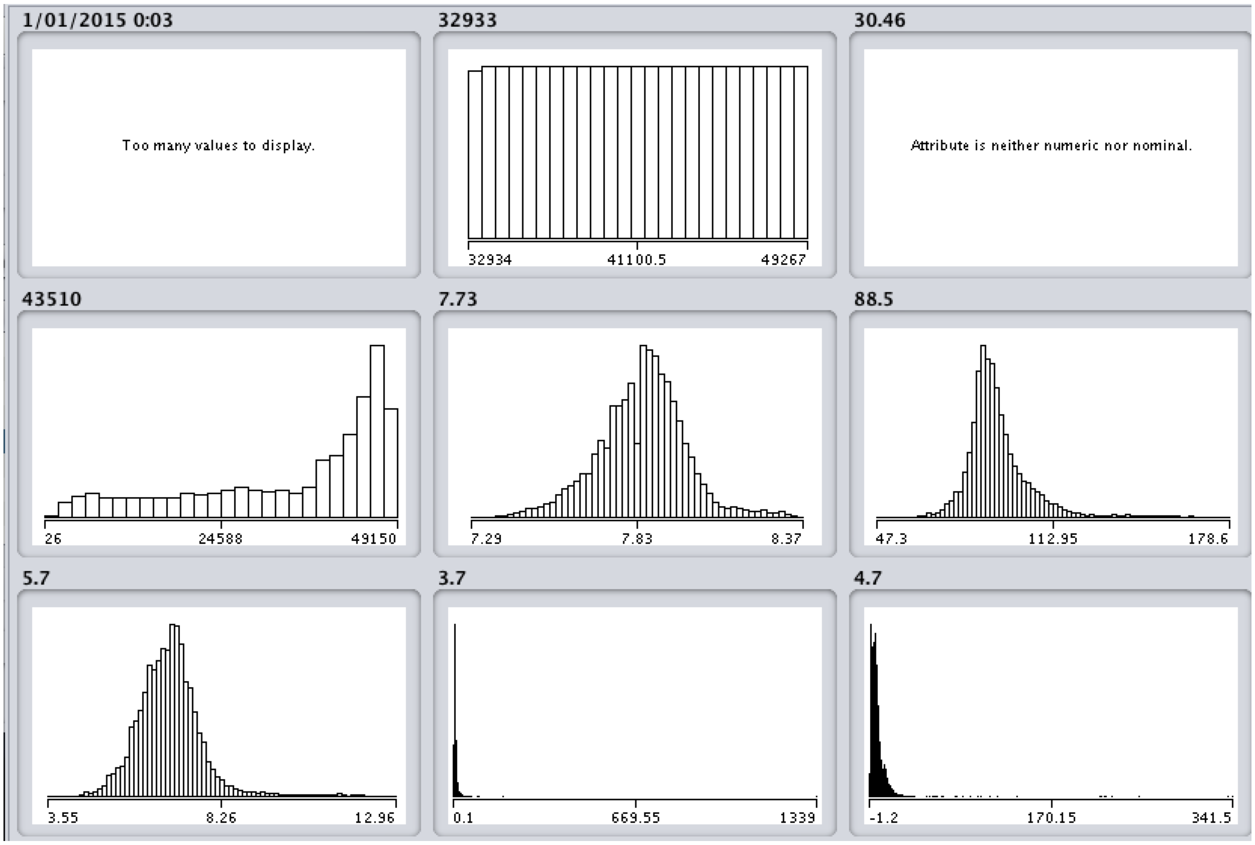
P1Q2.

This csv file contains nine set of variables, which is recorded from 0:03 of 1st January 2015 until 23:33 of 31st Dec 2015. It took data every half hours during all year and data index starts from 32933 and ends to 49267.




From the histogram graph, the fourth set demonstrated a right-skewed distribution from 26 to 49150 with mean 34670. The fifth, sixth and seventh set of data all demonstrate as a normal distribution, whose mean and standard deviation are 7.846 with 0.142, 92.423 with 12.206, 6.838 with 0.89 respectively. Whereas the last two graph are left-skewed distribution, which concentrate on minimum value 0.1 and -1.2 respectively.

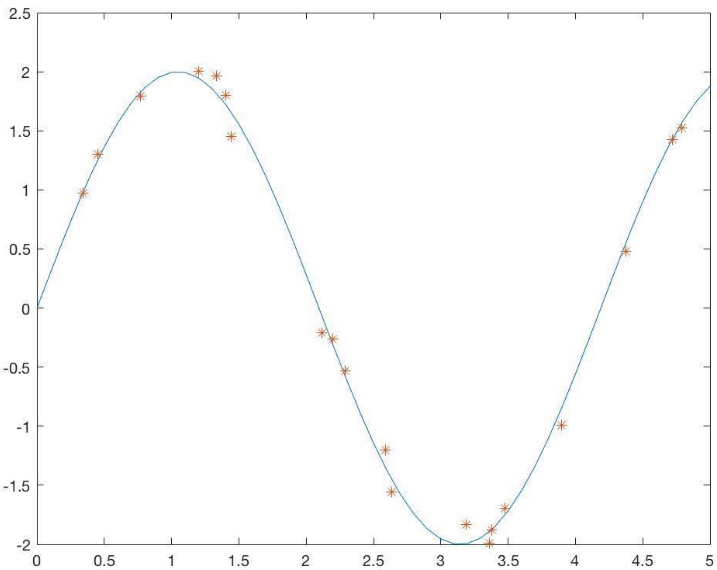
From the visualization part, we are trying to find the relationship between those data set.

- The dataset 1 (Timeline) and dataset 2 (Index) has a clearly positive linear distribution.
- The dataset 3 (30.46) and dataset 1 (Time)/dataset 2 (Index) are demonstrate a normal distribution. It initially increasing with time increasing, all the way until mid-year reach the maximum number, and then decreasing to the end of the year.
- The dataset 6 (88.5) and dataset 7 (5.7) are demonstrated to be a positive correlation. As visualizing shows, increasing number of dataset6 will results to an increasing number of dataset 7.

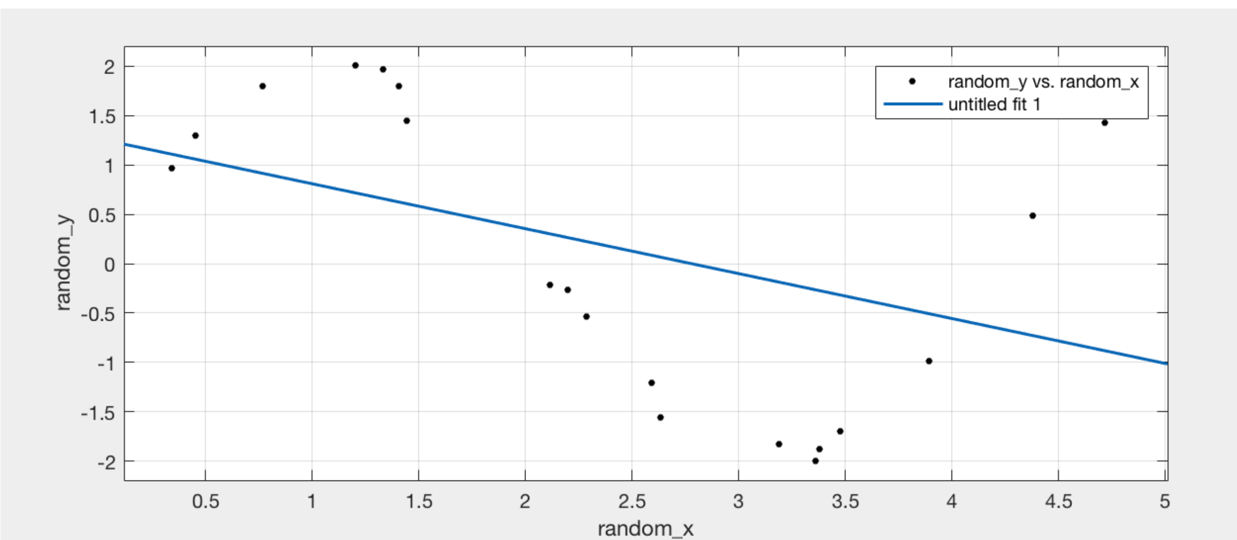
P1Q6.

```
function out = reverse(x, y)
  out = zeros(numel(x),1);
  if mod(numel(x), y) == 0
    loop_num = numel(x)/y ;
    loop_count = 0;
    while(loop_num > 0)
      for i = 1 : y
        out(i + loop_count*y) = x(numel(x) + i - (loop_count+1)*y);
      end
      loop_count = loop_count+1;
      loop_num = loop_num - 1;
    end
  else
    loop_num = uint8(mod(numel(x), y)) - 1;
    loop_count = 0;
    while(loop_num > 0)
      for i = 1 : y
        out(i + loop_count*y) = x(numel(x) + i - (loop_count+1)*y);
      end
      loop_count = loop_count+1;
      loop_num = loop_num - 1;
    end
    for i = 1:mod(numel(x),y)
      out(((loop_count)*y) + i) = x(i);
    end
  end
end
```

P2Q1.




(Figure: y=2sin(1.5x) and 20 random points with 0.1 variance)



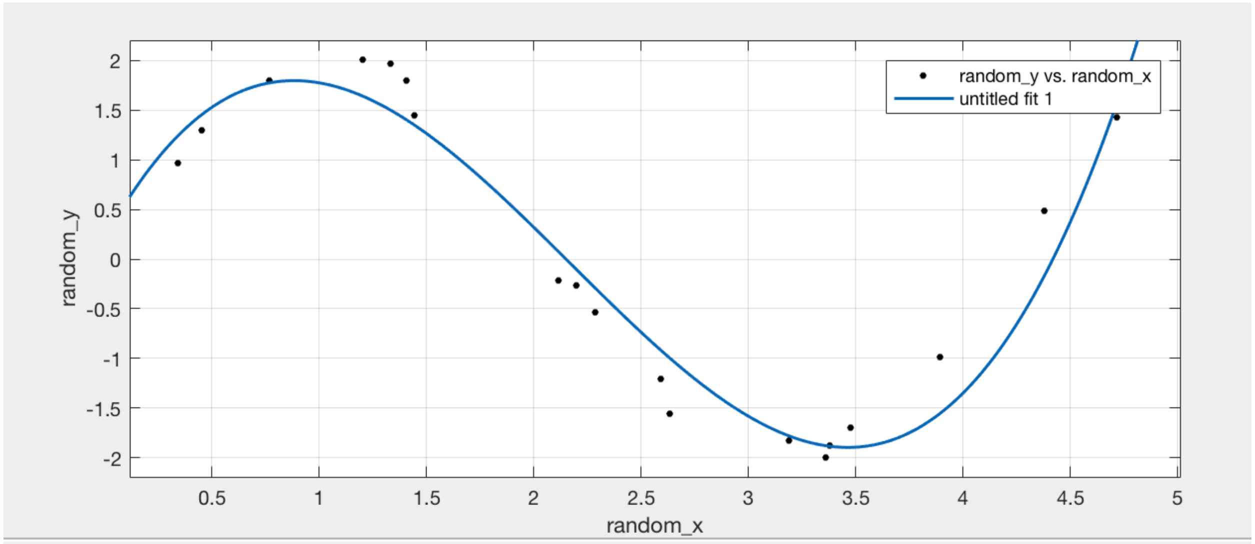

Data training with 1 degree:

SSE: 34.83

R-square 0.1765

Adjusted R-square: 0.1308

RMSE 1.391



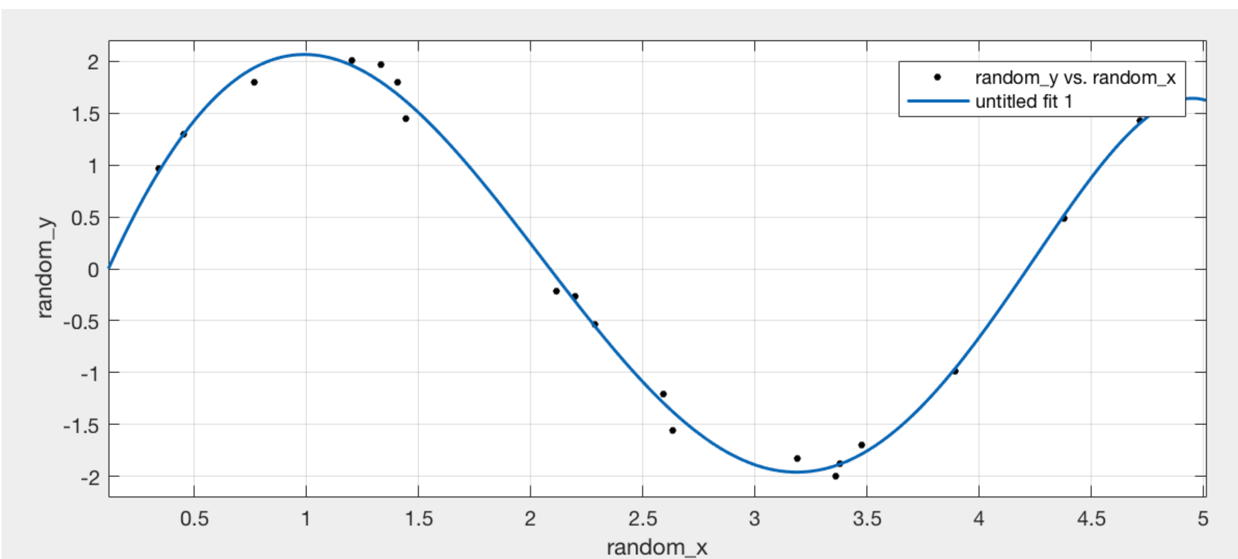

Data Training with Degree 3.

SSE: 2.255

R-square: 0.9467

Adjusted R-square: 0.9367

RMSE: 0.3754



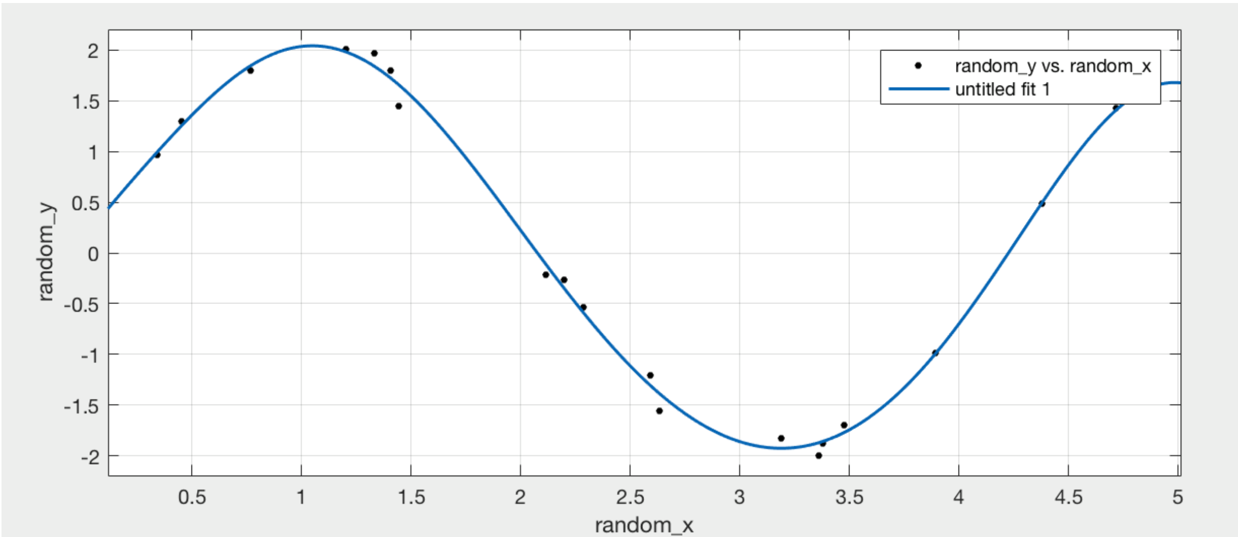

Data Training with Degree 5.

SSE: 0.1852

R-square: 0.9956

Adjusted R-square: 0.9941

RMSE: 0.115




Data Training with Degree 9.

SSE: 0.1571

R-square: 0.9963

Adjusted R-square: 0.9929

RMSE: 0.1254

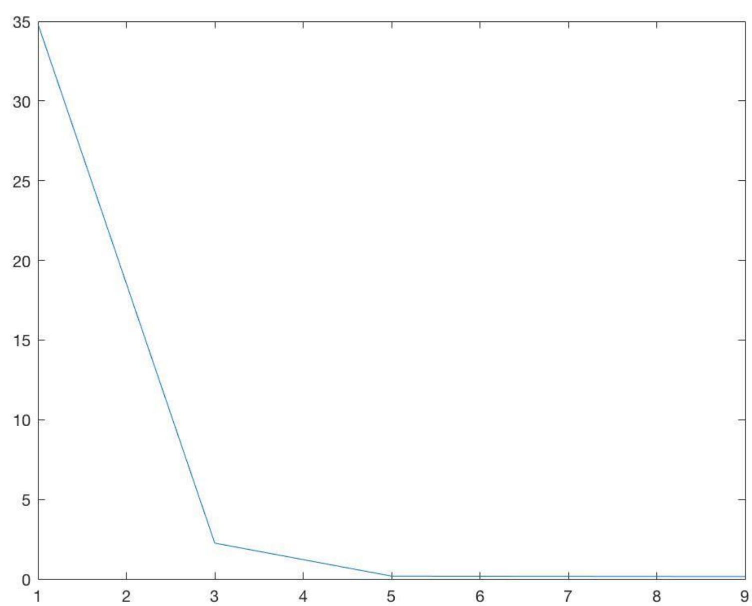
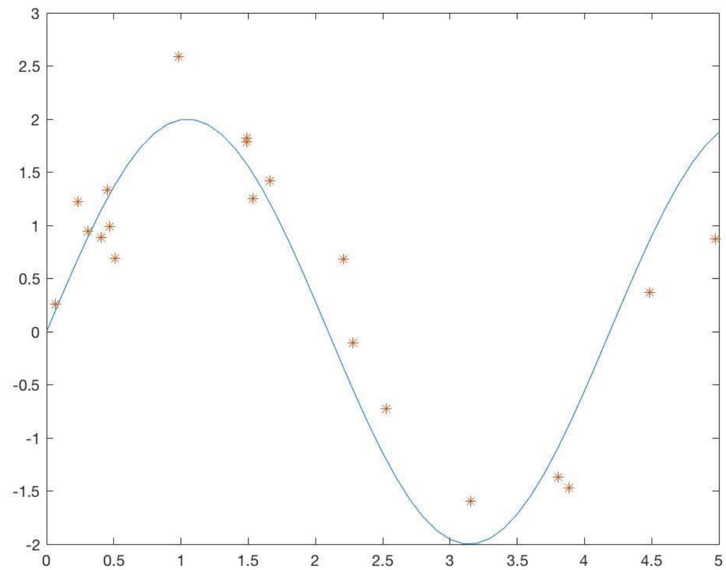
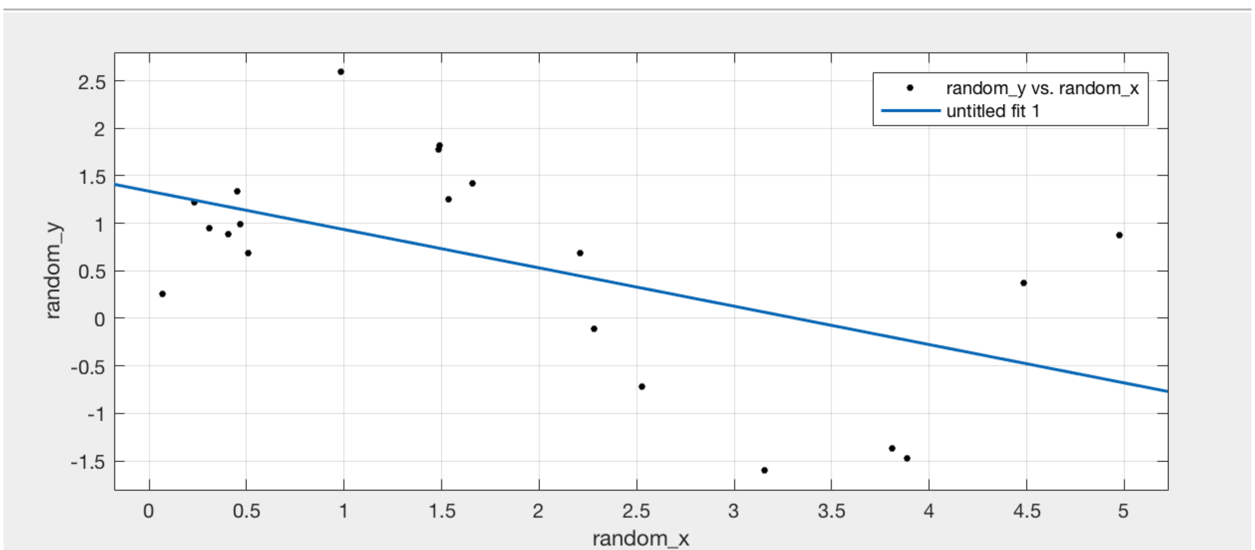



Figure: SSE Value vs Approximation Degree under 0.1 Variance




(Figure: y=2sin(1.5x) and 20 random points with 0.1 variance)



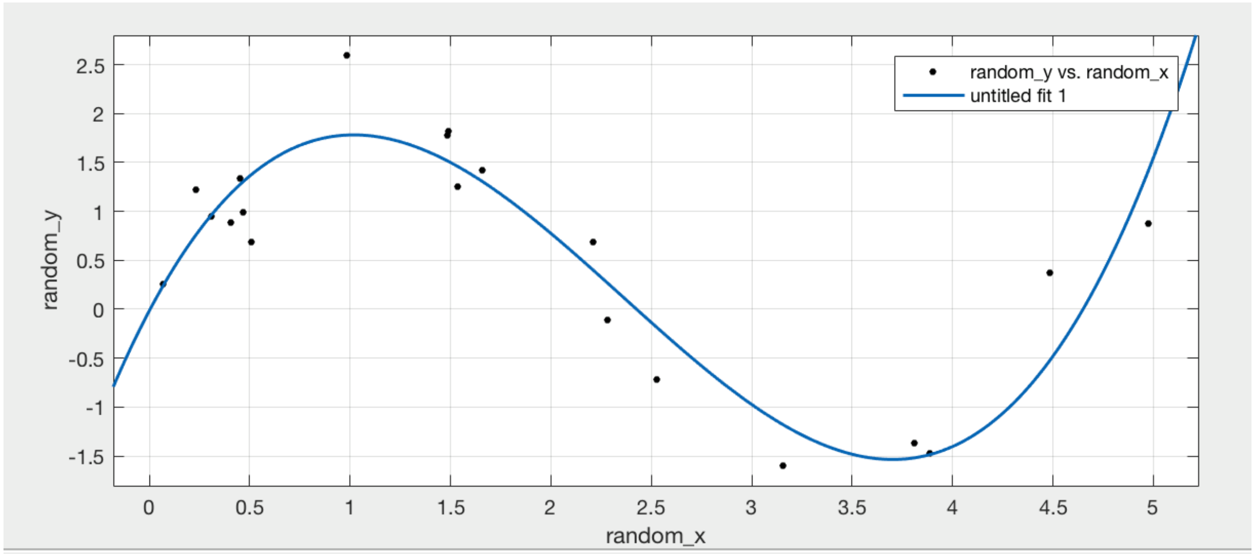

Data training with Degree 1

SSE: 17.52

R-Square: 0.2914

Adjusted R-square: 0.2521

RMSE: 0.9865



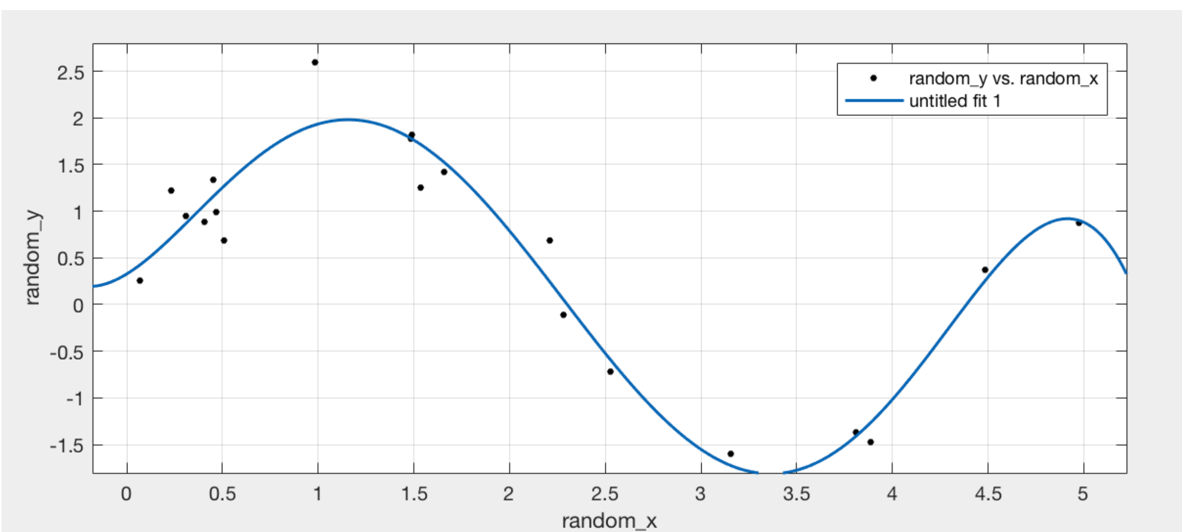

Data training with Degree 3

SSE: 3.558

R-Square: 0.8561

Adjusted R-Square: 0.8291

RMSE: 0.4715



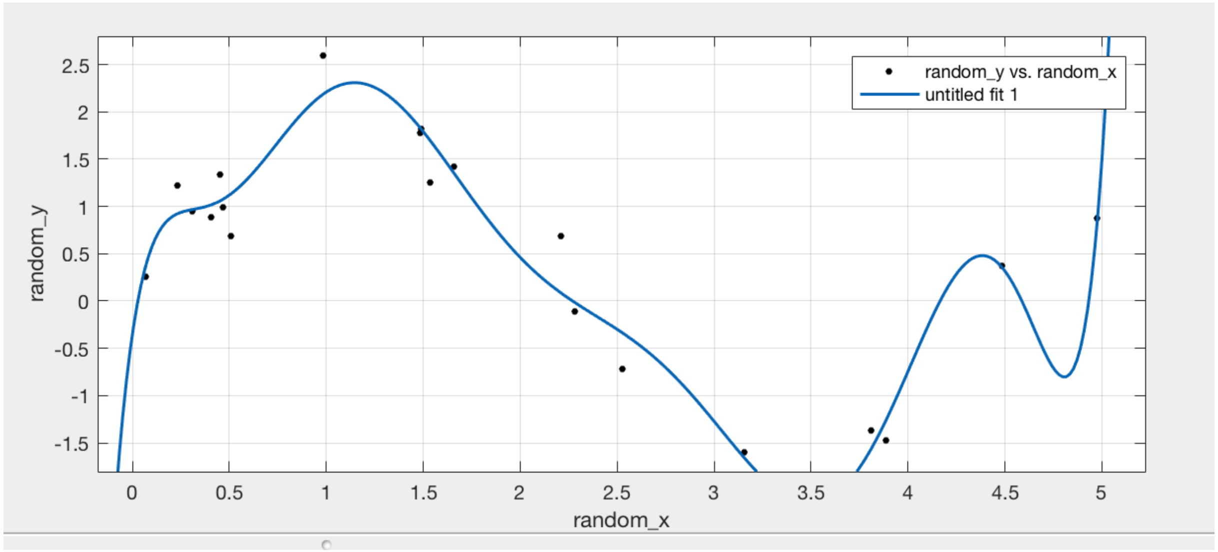

Data training with Degree 5

SSE: 1.708

R-Square: 0.9309

Adjusted R-Square: 0.9062

RMSE: 0.3493




Data training with Degree 9

SSE: 1.371

R-Square: 0.9445

Adjusted R-Square: 0.8946

RMSE: 0.3703

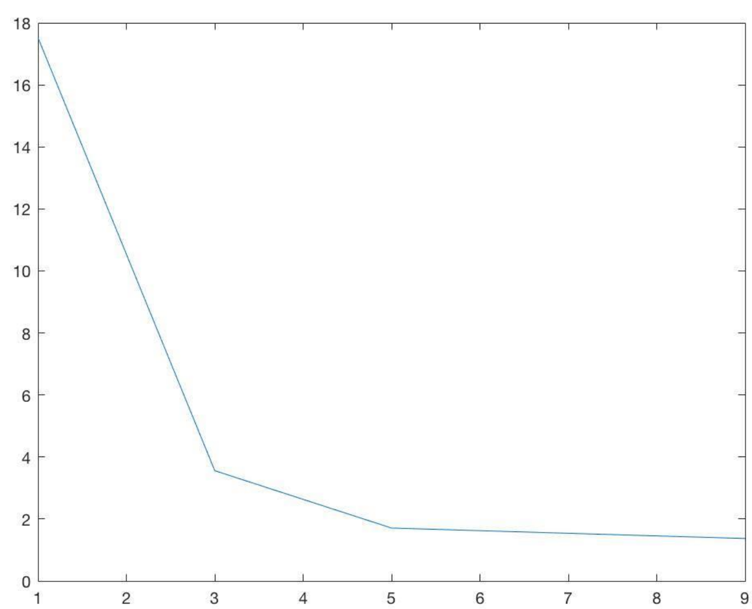
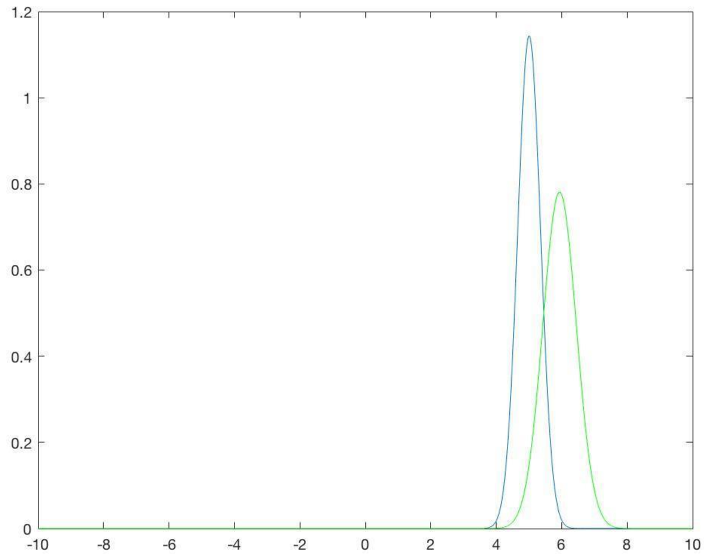



Figure: SSE Value vs Approximation Degree under 0.5 Variance

P2Q4.




.

Figure: Maximum Likelihoods of Reduced Iris data.

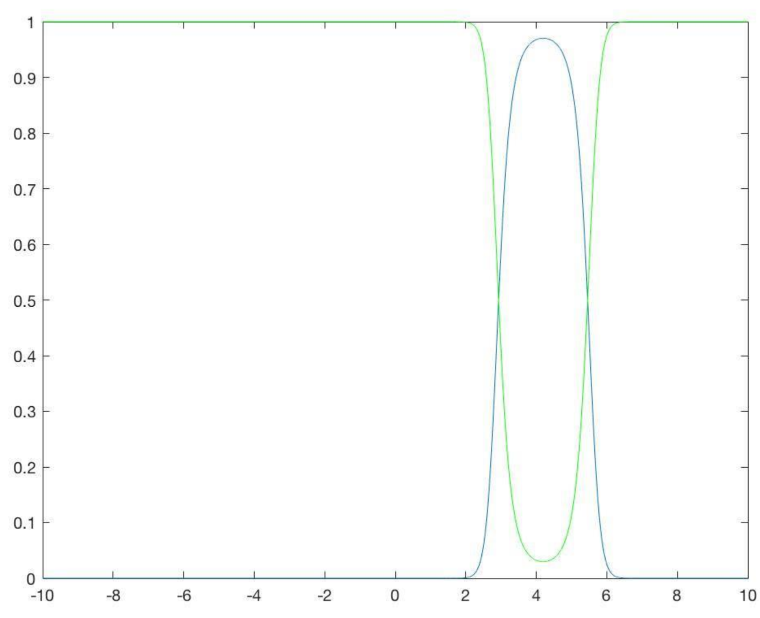



Figure: Class posteriors for Reduced Iris Data

Matlab Code:

```
iris = readtable('iris.txt', 'ReadVariableNames', false);
iris = table2cell(iris);

for i = 1:100
   if i <51
      x1(i) = iris(i,1);
   else
      x2(i - 50) = iris(i,1);
   end
end

x1 = x1.';
x2 = x2.';
x1 = cell2mat(x1);
x2 = cell2mat(x2);

mle1 = mle(x1);
mle2 = mle(x2);
x_array = [-10:.001:10];
figure(1);
y1 = normpdf(x_array, mle1(1), mle1(2));
plot(x_array, y1)
hold on
y2 = normpdf(x_array, mle2(1), mle2(2));
plot(x_array, y2, 'g')
hold off

figure(2);
```

```
for i = 1:20001
    y3(i) = y1(i)/(y1(i)+y2(i));
    y4(i) = y2(i)/(y1(i)+y2(i));
end
plot(x_array, y3)
hold on
plot(x_array, y4, 'g')
```

P3Q1

(a).

Running Matlab code, the number of error we get is 127. Therefore, training classification error is

Number of Errors / Number of Point = 127 / 500 = 25.4%

**Related Code:**

```
DataSet = readtable('pima_indians_diabetes.csv', 'ReadVariableNames', false);
DataSet = table2cell(DataSet);
counter1 = 1;
counter2 = 1;
for i = 1:500
  if (strcmp(DataSet(i,9), 'pos') == 1)
    c1(counter1,1:8) = DataSet(i,1:8);
    counter1 = counter1+1;
  else
    c2(counter2,1:8) = DataSet(i,1:8);
    counter2 = counter2+1;
  end
end
valid_interator = 1;
for i = 1:500
  valid(valid_interator, 1:8) = DataSet(i,1:8);
  valid_interator = valid_interator+1;
end
c1 = cell2mat(c1);
c2 = cell2mat(c2);
valid = cell2mat(valid);
cov1 = cov(c1);
cov2 = cov(c2);
mean1 = mean(c1);
```

```
mean2 = mean(c2);

y1 = mvnpdf(valid, mean1, cov1);

y2 = mvnpdf(valid, mean2, cov2);

error = 0;

predict_value = cell(500,1);

for i = 1:500

   if (y1(i) < y2(i))

      predict_value(i) = cellstr('neg');

   else

      predict_value(i) = cellstr('pos');

   end

   if (strcmp(predict_value(i),DataSet(i,9)) == 0)

      error = error+1;

   end

end
```

(b).

Running Matlab code for test classification, the number of error we get is 66. Therefore, test classification error is

Number of Errors / Number of Point = 66 / 268 = 24.6%

**Matlab Code**

```
DataSet = readtable('pima_indians_diabetes.csv', 'ReadVariableNames', false);

DataSet = table2cell(DataSet);

valid_iterator = 1;

for i = 501:768

   valid(valid_iterator, 1:8) = DataSet(i,1:8);

   valid_iterator = valid_iterator+1;

end

valid = cell2mat(valid);

z1 = mvnpdf(valid, mean1, cov1);
```

```
z2 = mvnpdf(valid, mean2, cov2);
error = 0;
predict_value = cell(268,1);
for i = 1:268
  if (z1(i) < z2(i))
    predict_value(i) = cellstr('neg');
  else
    predict_value(i) = cellstr('pos');
  end
  if (strcmp(predict_value(i),DataSet(i + 500,9)) == 0)
    error = error+1;
  end
end
```

(c).

Mean 1.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4.7802 | 140.4890 | 69.7253 | 21.7143 | 102.4286 | 35.3231 | 0.5672 | 36.2692 |

Mean 2.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3.2358 | 110.0283 | 67.9748 | 19.8711 | 67.5409 | 29.9921 | 0.4590 | 31.1855 |

Cov1.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 8.6603 | 11.4097 | 6.1763 | −3.9380 | −33.8882 | 1.1000 | −0.1219 | 19.1895 |
| 11.4097 | 806.5008 | 93.3982 | 1.7450 | 1.0654e+03 | 46.5349 | −1.2391 | 105.4868 |
| 6.1763 | 93.3982 | 324.8890 | 53.4794 | 152.3638 | 61.4594 | −1.2598 | 42.1466 |
| −3.9380 | 1.7450 | 53.4794 | 218.3146 | 611.6599 | 56.4678 | 0.0052 | −23.8814 |
| −33.8882 | 1.0654e+03 | 152.3638 | 611.6599 | 1.0826e+04 | 247.3216 | 7.4645 | −164.6149 |
| 1.1000 | 46.5349 | 61.4594 | 56.4678 | 247.3216 | 65.4896 | −0.3639 | 8.7567 |
| −0.1219 | −1.2391 | −1.2598 | 0.0052 | 7.4645 | −0.3639 | 0.2281 | −0.2483 |
| 19.1895 | 105.4868 | 42.1466 | −23.8814 | −164.6149 | 8.7567 | −0.2483 | 139.8740 |

Cov2.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 14.1503 | 7.6937 | 10.5415 | −4.5880 | −29.2644 | −3.5087 | −0.1273 | 18.3634 |
| 7.6937 | 968.3728 | 39.5992 | 2.7427 | 1.1351e+03 | 6.0058 | 0.7466 | 60.0168 |
| 10.5415 | 39.5992 | 494.0015 | 78.4515 | 270.3394 | 7.4423 | −0.1845 | 59.7263 |
| −4.5880 | 2.7427 | 78.4515 | 296.9234 | 1.2392e+03 | 35.2939 | 1.7236 | −30.5470 |
| −29.2644 | 1.1351e+03 | 270.3394 | 1.2392e+03 | 1.9763e+04 | 30.7702 | 5.3007 | 162.4475 |
| −3.5087 | 6.0058 | 7.4423 | 35.2939 | 30.7702 | 56.3220 | 0.3371 | −17.0913 |
| −0.1273 | 0.7466 | −0.1845 | 1.7236 | 5.3007 | 0.3371 | 0.1555 | −0.3581 |
| 18.3634 | 60.0168 | 59.7263 | −30.5470 | 162.4475 | −17.0913 | −0.3581 | 114.5514 |

P3Q2.

**P3Q5.**

**M and H1 Value:**

-169.769126255560 -169.769126255560 -153.987061905532 -153.987061905532 -
134.671376212291 -109.769126255560 Inf -14.6713762122905 Inf -
74.6713762122905 -109.769126255560 -153.987061905532 -134.671376212291 -
169.769126255560 -109.769126255560 -109.769126255560 -74.6713762122905 Inf-
14.6713762122905 -14.6713762122905

**M and K1 Value:**

| | | | |
|---|---|---|---|
| 694.932323934640 | 662.732703004150 | 631.621908886040 | 601.596535275394 |
| 572.653049197447 | 544.787786823609 | 517.996948880953 | 492.276595495771 |
| 467.622640256742 | 444.030843211791 | 421.496802420375 | 400.015943563853 |
| 379.583506963508 | 360.194531159384 | 341.843831951819 | 324.525975486758 |
| 308.235243557651 | 292.965588778855 | 278.710576631234 | 265.463310558606 |
| 253.216335267740 | 241.961512115841 | 231.689858919299 | 222.391344655224 |
| 214.054627343051 | 206.666720921067 | 200.212574287935 | 194.674543120838 |
| 190.031733105538 | 186.259192689460 | 183.326935829359 | 181.198782674900 |
| 179.831021952489 | 179.170927351365 | 179.155206650030 | 179.708531501912 |
| 180.742389522574 | 182.154612391893 | 183.830042752956 | 185.642865017160 |
| 187.461073425890 | 189.153308990186 | 190.597821450905 | 191.692652337506 |
| 192.365478282509 | 192.581196960218 | 192.345548734539 | 191.703892562162 |
| 190.735427999510 | 189.544211278481 | 188.248855405650 | 186.972716712011 |
| 185.835828611042 | 184.949152576538 | 184.411131260420 | 184.306167523733 |
| 184.704512294726 | 185.663057529421 | 187.226623027147 | 189.429440923842 |
| 192.296647156948 | 195.845672012320 | 200.087480427544 | 205.027650707899 |
| 210.667302848118 | 217.003899481812 | 224.031947271700 | 231.743626861874 |
| 240.129376977209 | 249.178453864154 | 258.879481612308 | 269.221002376203 |
| 280.192028495233 | 291.782591418975 | 303.984275732635 | 316.790721091659 |
| 330.198071192257 | 344.205347611058 | 358.814727797436 | 374.031710701231 |
| 389.865160061081 | 406.327223474907 | 423.433133971073 | 441.200908766725 |

| | | | |
|---|---|---|---|
| 459.650966241466 | 478.805686180156 | 498.688939763663 | 519.325614736672 |
| 540.741158112279 | 562.961154338331 | 586.010951759107 | 609.915345100391 |
| 634.698317081999 | 660.382838435870 | 686.990722722914 | 714.542530399731 |
| 743.057515490450 | 772.553607817864 | 803.047423870824 | 834.554299867354 |

**M and K2 Value**

682.398756308608 636.029981737874 592.011872283877 550.321605057753 510.934125197960
473.822002250326 438.955295775757 406.301434875496 375.825116902666 347.488231217678
321.249814485490 297.066044766966 274.890282589117 254.673168316788 236.362786411479
219.904908272007 205.243325725184 192.320285919873 181.077034073381 171.454461765110
163.393844059325 156.837628322031 151.730212373865 148.018622781879 145.652980651262
144.586627938675 144.775786318725 146.178633015037 148.753697034641 152.457490261243
157.241269391662 163.046752757879 169.800473226423 177.406241184336 185.734984166711
194.611204078624 203.795841702185 212.967145417325 221.705055962300 229.490822413291
235.739437626780 239.880169709137 241.479787175608 240.366597055864 236.691078191613
230.885165148683 223.543807053551 215.291976328774 206.688464918496 198.180915465286
190.100159522958 182.675373515471 176.056102339778 170.333680252830 165.559158728495
161.757214536351 158.936429487834 157.096539765476 156.233197322903 156.340687776887
157.412989267352 159.443543215386 162.424116985576 166.343139089153 171.183856170174
176.922588559044 183.527254379289 190.956211575168 199.157364266154 208.067433625515
217.611344792461 227.701860417412 238.239896691160 249.116326124965 260.216351201536
271.427485335865 282.651550002214 293.819801099163 304.908635394184 315.952062987999
327.047166487992 338.350474896097 350.065927181620 362.427542970892 375.680930513383
390.067162438721 405.811011585625 423.113950676746 442.151240588265 463.071973600700
486.000936156639 511.041382660536 538.278095267913 567.780352802729 599.604614209090
633.796840420339 670.394447667143 709.427920843689 750.922130424768 794.897399579461

**Matlab Code**

```
x = [randn(30,1); 5+randn(30,1)];

H = hist(x, 20);

figure(2);

[K1, xi1, bw] = ksdensity(x);

plot(xi1, K1,'--');

hold on;

[K2, xi2] = ksdensity(x, 'width', bw/2);

plot(xi2, K2)


M = normpdf(x)

Klmh = sum(M.*log2(M)-log2(H))

Klmk1 = sum(M.*log2(M)-log2(K1))

Klmk2 = sum(M.*log2(M)-log2(K2))
```