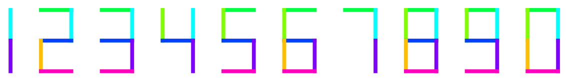
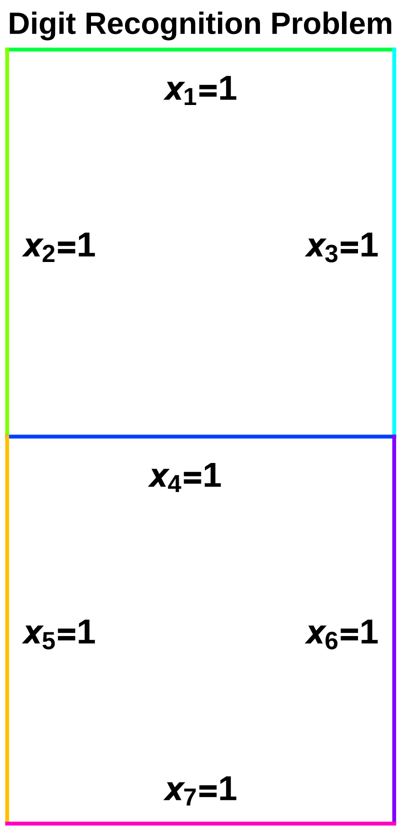
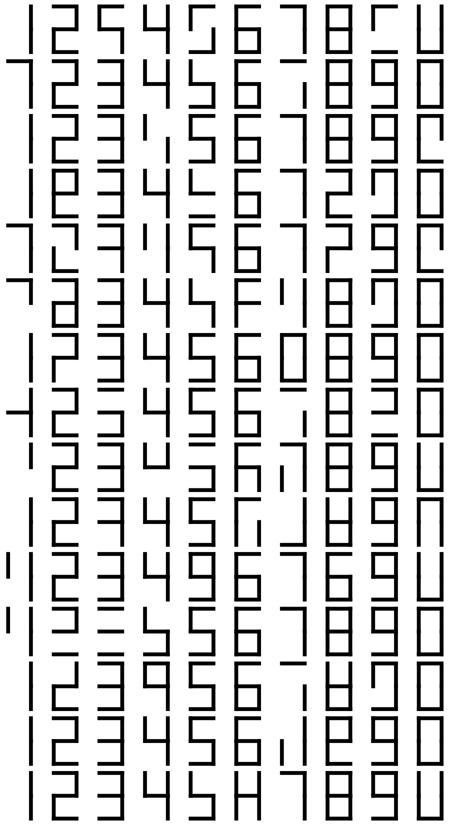
Digit Recognition Example

Breiman et al. (1984, §2.6.1, p.43). The digits 1, 2, 3, ..., 9, 0 represented by line segments. The are 7 line segments as illustrated for each digit. So the digit 8 corresponds to $x_1 =. .. = x_7 = 1$. Each line segment has 10% probability of flipping on or off.





Below is a random sample. Each row corresponds to the digits 1, 2, ..., 9, 0 but with some line segments flipped.



The table below shows the settings for $x_1, \ldots, x_7$ corresponding to digits 1, 2, ..., 9, 0.

digit	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	0	0	1	0	0	1	0
2	1	0	1	1	1	0	1
3	1	0	1	1	0	1	1
4	0	1	1	1	0	1	0
5	1	1	0	1	0	1	1
6	1	1	0	1	1	1	1
7	1	0	1	0	0	1	0
8	1	1	1	1	1	1	1
9	1	1	1	1	0	1	1
0	1	1	1	0	1	1	1

Bayes error rate BFOS Digit Recognition

Derivation of Bayes Solution

Let $\mathcal{D}$ be the true digit and we assme the prior distribution is uniform, $\Pr\{\mathcal{D} = d\} = 1/10$ for $d = 1, 2, \ldots, 9, 0$. The electronic display corresponds to setting $x_i = 0/1$ for $i = 1, \ldots, 7$.

In the table below we show for each digit, $d = 1, 2, \ldots, 9, 0$, (in blue) the corresponding indicies, i, for the x's. For example the first row shows that for $d = 1$, we need to set $x_3 = x_6 = 1$ and all the other x's are zero. The settings for the x's correspond to the binary representations of the digits 18, 93, …, 119 as shown. All possible settings for the x's correspond to the binary representations for $X = 0, \ldots, 127$.

1: 18	3	6					
2: 93	1	3	4	5	7		
3: 91	1	3	4	6	7		
4: 58	2	3	4	6			
5: 107	1	2	4	6	7		
6: 111	1	2	4	5	6	7	
7: 82	1	3	6				
8: 127	1	2	3	4	5	6	7
9: 123	1	2	3	4	6	7	
0: 119	1	2	3	5	6	7	

Consider $X = 25$, the binary representation is,

```
In[1]:= IntegerDigits[25, 2, 7]
Out[1]= {0, 0, 1, 1, 0, 0, 1}
```

So $X = 25$ corresponds to $x_3 = x_4 = x_7 = 1$ and $x_1 = x_2 = x_5 = x_6 = 0$. Now suppose that $D = 4$ which corresponds to $x_2 = x_3 = x_4 = x_6 = 1$ and $x_1 = x_5 = x_7 = 0$. So

$\Pr\{X = 25 \mid D = 4\} = \Pr\{x_3, x_4 \text{ stay on}\} \times \Pr\{x_2, x_6 \text{ turn off}\} \times \Pr\{x_7 \text{ on}\} \times \Pr\{\text{rest stay off}\}$

Assume the probability of switching from off to on is α and from on to off is $1 - \alpha$. We take $\alpha = 0.1$ for illustration. Then

$\Pr\{x_3, x_4 \text{ stay on}\} = 0.9^2$

$\Pr\{x_2, x_6 \text{ turn off}\} = 0.1^2$

$\Pr\{x_7 \text{ switches on}\} = 0.1$

$\Pr\{\text{rest stay off}\} = 0.9^2$

Hence $\Pr\{X = 25 \mid D = 4\} = 0.9^4 \times 0.1^3 = 0.0006561$.

Assuming a uniform prior, $\Pr\{X = 25, D = 4\} = 0.9^4 \times 0.1^3 = 0.00006561$.

The following *Mathematica* code computes these probabilities.

```
In[2]:= ProbOn = 0.1;
ProbOff = 1 - ProbOn;
NumLines = {(*1*){3, 6}, (*2*){1, 3, 4, 5, 7}, (*3*){1, 3, 4, 6, 7}, (*4*)
   {2, 3, 4, 6}, (*5*){1, 2, 4, 6, 7}, (*6*){1, 2, 4, 5, 6, 7}, (*7*){1, 3, 6},
   (*8*){1, 2, 3, 4, 5, 6, 7}, (*9*){1, 2, 3, 4, 6, 7}, (*0*){1, 2, 3, 5, 6, 7}};
pdx = Table[
   p = Array[ProbOn &, 7];
   p[[NumLines[[id]]]] = ProbOff;
   k = IntegerDigits[ix - 1, 2, 7];
   kp = Transpose[{k, p}];
   prob = Times @@ (If[First[#] == 0, 1 - Last[#], Last[#]] &) /@ kp;
   {If[id ≠ 10, id, 0], ix - 1, prob / 10}, (*digit,X,prob*)
   {id, 10}, {ix, 2^7}];
```

The result is in the array **pdx** of dimensions:

```
In[6]:= Dimensions[pdx]
Out[6]= {10, 128, 3}
```

The first dimension corresponds to $d = 1, \ldots, 9, 0$; the second dimension corresponds to

$X = 0, 1, \ldots, 127$ and the third contains $\{d, x, \Pr\{X = x \mid D = d\}\}$. For example,

```
In[7]:= pdx[[4, 26]]
Out[7]= {4, 25, 0.00006561}

In[8]:= pdx[[3, 92]]
Out[8]= {3, 91, 0.0478297}

In[9]:= pdx[[7, 92]]
Out[9]= {7, 91, 0.00059049}

In[10]:= Select[pdx[[3]], Last[#] == Last[pdx[[3, 92]]] &]
Out[10]= {{3, 91, 0.0478297}}

In[11]:= pxd = Transpose[pdx, {2, 1, 3}];

In[12]:= pxd[[92]]
Out[12]= {{1, 91, 0.00006561}, {2, 91, 0.00059049}, {3, 91, 0.0478297}, {4, 91, 0.00006561},
  {5, 91, 0.00059049}, {6, 91, 0.00006561}, {7, 91, 0.00059049},
  {8, 91, 0.00059049}, {9, 91, 0.00531441}, {0, 91, 0.00006561}}

In[13]:= pxd = Transpose[pdx, {2, 1, 3}];
t = Table[
   pd = Last[Transpose[pxd[[i]]]];
   pd = pd / Total[pd];
   M = Join[Transpose[pxd[[i]]], {pd}] // Transpose;
   Select[M, Last[#] == Max[pd] &],
   {i, 1, 128}];
delta = Map[First[Take[#, 2]] &, t, {2}];

In[16]:= delta[[91]]
Out[16]= {3, 7}

In[17]:= Dimensions[pxd]
Out[17]= {128, 10, 3}
```

Bayes Solution Computations

Definitions and Table

```
In[18]:= NumLines = {(*1*){3, 6}, (*2*){1, 3, 4, 5, 7}, (*3*){1, 3, 4, 6, 7}, (*4*)
   {2, 3, 4, 6}, (*5*){1, 2, 4, 6, 7}, (*6*){1, 2, 4, 5, 6, 7}, (*7*){1, 3, 6},
   (*8*){1, 2, 3, 4, 5, 6, 7}, (*9*){1, 2, 3, 4, 6, 7}, (*0*){1, 2, 3, 5, 6, 7}};

In[19]:= tobin[x_] := Module[{t = Array[0 &, 7]}, t[[x]] = 1; t];
dcodes = FromDigits[#, 2] & /@ (tobin /@ NumLines);
TableForm[{Append[Range[9], 0], dcodes}, TableHeadings → {{"digits", "X"}, None}]
```

Out[21]//TableForm=

digits	1	2	3	4	5	6	7	8	9	0
X	18	93	91	58	107	111	82	127	123	119

In[22]:=
```
s1 = ToString /@ Append[Range[9], 0];
```

In[23]:=
```
s2 = StringJoin[": ", #] & /@ (ToString /@ dcodes);
```

In[24]:=
```
rh = Style[#, FontFamily → "Helvetica", FontSize → 14, FontColor → Blue] & /@
    MapThread[StringJoin, {s1, s2}];
```

In[25]:=
```
TableForm[Map[Style[#, FontColor → Black, FontWeight → Bold, FontSize → 13] &,
   NumLines, {2}], TableHeadings → {rh, None}]
```

Out[25]//TableForm=

1: 18	3	6					
2: 93	1	3	4	5	7		
3: 91	1	3	4	6	7		
4: 58	2	3	4	6			
5: 107	1	2	4	6	7		
6: 111	1	2	4	5	6	7	
7: 82	1	3	6				
8: 127	1	2	3	4	5	6	7
9: 123	1	2	3	4	6	7	
0: 119	1	2	3	5	6	7	

Conditional Distribution

For each

$\Pr\{X \mid \mathcal{D}\}$ is the probability

Here we find $\Pr\{X \mid \mathcal{D}\}$. First for the case $\mathcal{D} = 1, 2, \ldots, 9, 0$. $\Pr\{X \mid \mathcal{D}\} \propto \Pr\{X\}$. In fact, we can write $\Pr\{X\} = \Pr\{X \mid \mathcal{D}\}/10$. For Bayes rule we need $\Pr\{\mathcal{D} \mid X\}$.

In[26]:=
```
ProbOn = 0.1;
ProbOff = 1 - ProbOn;
pdx = Table[
   p = Array[ProbOn &, 7];
   p[[NumLines[[id]]]] = ProbOff;
   k = IntegerDigits[ix - 1, 2, 7];
   kp = Transpose[{k, p}];
   prob = Times @@ (If[First[#] == 0, 1 - Last[#], Last[#]] &) /@ kp;
   {If[id ≠ 10, id, 0], ix - 1, prob / 10}, (*digit,X,prob*)
   {id, 10}, {ix, 2^7}];
```

We see that the for id = 10 which corresponds to the digit "0", the maximum joint probability for id, ix is given by,

In[29]:=
```
pxd = Transpose[pdx, {2, 1, 3}];
t = Table[
   pd = Last[Transpose[pxd[[i]]]];
   pd = pd / Total[pd];
   M = Join[Transpose[pxd[[i]]], {pd}] // Transpose;
   Select[M, Last[#] == Max[pd] &],
   {i, 1, 128}];
delta = Map[First[Take[#, 2]] &, t, {2}];
```

Each value in the table shows for a given ix = 1, …, 128 the corresponding entries:

d, x, $p_{x,d}$, $P\{D = d \mid X = x\}$

Let $p_{x,d} = \Pr\{X = x, \mathcal{D} = d\}$. Then the Bayes error rate, $\in\{\mathcal{D} \neq \delta(\mathcal{X})\}$. Assuming no ties,

$$\sum_{x,d} p_{x,d}\, \mathcal{I}(\mathcal{D} \neq \delta(X))$$

Random choice is equivalent to letting $\delta(X)$ = set of optimal choices for $\mathcal{D}$.

$$\eta = 1 - \sum_{x,d} p_{x,d}\, \mathcal{I}(\mathcal{D} \in \delta(X)) / \left| \delta(X) \right|$$

In[32]:=
```
1 - Sum[
  dHat = delta[[ix]];
  p = Last[pdx[[id, ix]]] ;
  d = If[id == 10, 0, id];
  J = If[MemberQ[dHat, d], 1.0 / Length[dHat], 0.0];
  p * J,
  {id, 10}, {ix, 128}]
```

Out[32]= 0.259978

Solution for $\alpha = 0.01, 0.02, \ ..., \ 0.2$

```
In[33]:= B = Table[
   NumLines = {(*1*){3, 6}, (*2*){1, 3, 4, 5, 7}, (*3*)
     {1, 3, 4, 6, 7}, (*4*){2, 3, 4, 6}, (*5*){1, 2, 4, 6, 7}, (*6*)
     {1, 2, 4, 5, 6, 7}, (*7*){1, 3, 6}, (*8*){1, 2, 3, 4, 5, 6, 7},
     (*9*){1, 2, 3, 4, 6, 7}, (*0*){1, 2, 3, 5, 6, 7}};
  tobin[x_] := Module[{t = Array[0 &, 7]}, t[[x]] = 1; t];
  dcodes = FromDigits[#, 2] & /@ (tobin /@ NumLines);
  ProbOn = alpha;
   ProbOff = 1 - ProbOn;
   NumLines = {(*1*){3, 6}, (*2*){1, 3, 4, 5, 7}, (*3*)
     {1, 3, 4, 6, 7}, (*4*){2, 3, 4, 6}, (*5*){1, 2, 4, 6, 7}, (*6*)
     {1, 2, 4, 5, 6, 7}, (*7*){1, 3, 6}, (*8*){1, 2, 3, 4, 5, 6, 7},
     (*9*){1, 2, 3, 4, 6, 7}, (*0*){1, 2, 3, 5, 6, 7}};
   pdx = Table[
     p = Array[ProbOn &, 7];
     p[[NumLines[[id]]]] = ProbOff;
     k = IntegerDigits[ix - 1, 2, 7];
     kp = Transpose[{k, p}];
     prob = Times @@ (If[First[#] == 0, 1 - Last[#], Last[#]] &) /@ kp;
     {If[id ≠ 10, id, 0], ix - 1, prob / 10}, (*digit,X,prob*)
     {id, 10}, {ix, 2^7}];
  pxd = Transpose[pdx, {2, 1, 3}];
   px = Table[
     pXi = Last[Transpose[pxd[[i]]]];
     pXi = pXi / Total[pXi];
     pM = Join[Transpose[pxd[[i]]], {pXi}] // Transpose;
     Select[pM, Last[#] == Max[pXi] &],
     {i, 1, 128}];
  delta = Map[First[Take[#, 2]] &, px, {2}];
  {alpha, 1 - Sum[
       dHat = delta[[ix]];
       p = Last[pdx[[id, ix]]] ;
       d = If[id == 10, 0, id];
       J = If[MemberQ[dHat, d], 1.0 / Length[dHat], 0.0];
       p * J,
       {id, 10}, {ix, 128}]}
   , {alpha, 0.01, 0.21, 0.01}]; TableForm[B, TableHeadings → {None, {"α", "η"}}]
```

Out[33]//TableForm=

α	η
0.01	0.0269355
0.02	0.0537248
0.03	0.0803424
0.04	0.106764
0.05	0.132968
0.06	0.158931
0.07	0.184632
0.08	0.210053
0.09	0.235174
0.1	0.259978
0.11	0.284448
0.12	0.308569
0.13	0.332326
0.14	0.355706
0.15	0.378697
0.16	0.401286
0.17	0.423462
0.18	0.445217
0.19	0.466541
0.2	0.487427
0.21	0.507866

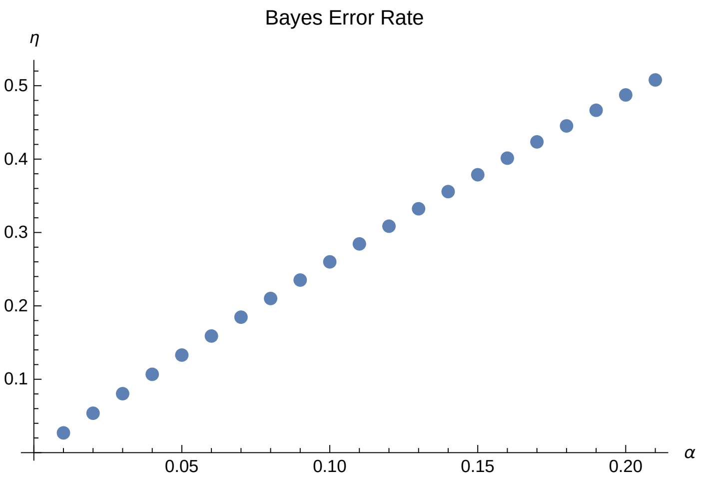
Remark: The Bayes error rate can not be worse than random guessing so we need to restrict $\alpha < 0.2$. We should check what happens when alpha=0.2 to classification algorithms such as C5 and RF.

In[34]:=
```
B2 = Most[B];
B2b =
  StringJoin[ToString[NumberForm[Round[100 * Last[#], 0.1], {5, 1}]], "%"] & /@ B2;
B2a = ToString[NumberForm[First[#], {4, 2}]] & /@ B2;
B3 = Partition[Transpose[{B2a, B2b}], 7];
f[x_] := Style[x, Blue, Bold];
TableForm[Partition[Flatten[Transpose[B3, {2, 1, 3}]], 4],
 TableHeadings → {None, {"α" // f, "η" // f, "α" // f, "η" // f}}]
```

Out[39]//TableForm=

α	η	α	η
0.01	2.7%	0.08	21.0%
0.02	5.4%	0.09	23.5%
0.03	8.0%	0.10	26.0%
0.04	10.7%	0.11	28.4%
0.05	13.3%	0.12	30.9%
0.06	15.9%	0.13	33.2%
0.07	18.5%	0.14	35.6%

In[40]:=
```
ListPlot[B, PlotLabel → "Bayes Error Rate",
 AxesLabel → {"α", "η"}, PlotStyle → {PointSize[Large]}]
```

Out[40]=


From the above plot we see that the relationship between the Bayes error rate and α looks linear. Indeed fitting a straight line we find $\eta = 0.0113405 + 2.4316\,\alpha$ with $R^2 = 99.9\%$ so the fit is essentially perfect. Note that the intercept term is highly significant even though we know that when $\alpha = 0$ then $\eta = 0$.

Further Comments

By computing all solutions for $\alpha = 0.01, \ldots, 0.49$ in steps of 0.01, it was found that $\delta(X)$ for
$X \in \{18, 93, 91, 58, 107, 111, 82, 127, 123, 119\}$
is unique and corresponds to the expected digit.

When $\alpha = 0.5$, $\Pr\{D = d \mid X\} = 0.1$ for all d and X, that is for any observed X all values of d are equally probable.

Predict η given α

In[41]:=
```
?LinearModelFit
```

LinearModelFit[$\{y_1, y_2, \ldots\}, \{f_1, f_2, \ldots\}, x$] constructs a linear model of the form $\beta_0 + \beta_1 f_1 + \beta_2 f_2 + \ldots$ that fits the y_i for successive x values 1, 2, ….

LinearModelFit[$\{\{x_{11}, x_{12}, \ldots, y_1\}, \{x_{21}, x_{22}, \ldots, y_2\}, \ldots\}, \{f_1, f_2, \ldots\}, \{x_1, x_2, \ldots\}$] constructs a linear model of the form $\beta_0 + \beta_1 f_1 + \beta_2 f_2 + \ldots$ where the f_i depend on the variables x_k.

LinearModelFit[$\{m, v\}$] constructs a linear model from the design matrix m and response vector v. ≫

In[42]:=
```
ans = LinearModelFit[B, α, α]
```

Out[42]= FittedModel[$0.0126845 + 2.41328\,\alpha$]

Evaluating the regression at $\alpha = 0.1$ and $\alpha = 0$.

In[43]:=
```
{ans[0.0], ans[0.01], ans[0.1], ans[0.2]}
```

Out[43]= {0.0126845, 0.0368173, 0.254012, 0.49534}

In[44]:=
```
ans[0.0]
```
Out[44]= 0.0126845

In[45]:=
```
ans["BestFitParameters"]
```
Out[45]= {0.0126845, 2.41328}

In[46]:=
```
ans["ParameterTable"]
```
Out[46]=

	Estimate	Standard Error	t-Statistic	P-Value
1	0.0126845	0.00266279	4.76361	0.000135089
α	2.41328	0.0212064	113.8	2.15072×10^{-28}

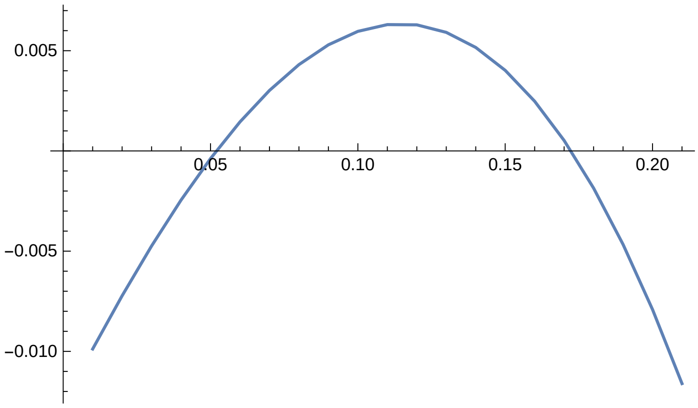
In[47]:=
```
ans["ANOVATable"]
```
Out[47]=

	DF	SS	MS	F-Statistic	P-Value
α	1	0.448441	0.448441	12950.3	2.15072×10^{-28}
Error	19	0.000657928	0.0000346278		
Total	20	0.449099			

In[48]:=
```
ans["RSquared"]
```
Out[48]= 0.998535

There is clear lack of fit. We need to add a quadratic term!

In[49]:=
```
ListPlot[{First /@ B, ans["FitResiduals"]} // Transpose, Joined → True]
```
Out[49]=



Quadratic

In[50]:=
```
ans = LinearModelFit[B, {α, α^2}, α]
```
Out[50]= FittedModel[$-0.0017144 + 2.7889\,\alpha - 1.70738\,\alpha^2$]

In[51]:=
```
ans["ParameterTable"]
```
Out[51]=

	Estimate	Standard Error	t-Statistic	P-Value
1	−0.0017144	0.0003398	−5.04531	0.0000841775
α	2.7889	0.00711449	392.003	7.69287×10^{-37}
α^2	−1.70738	0.0314072	−54.3627	2.03123×10^{-21}

In[52]:=
```
ans["ANOVATable"]
```

Out[52]=

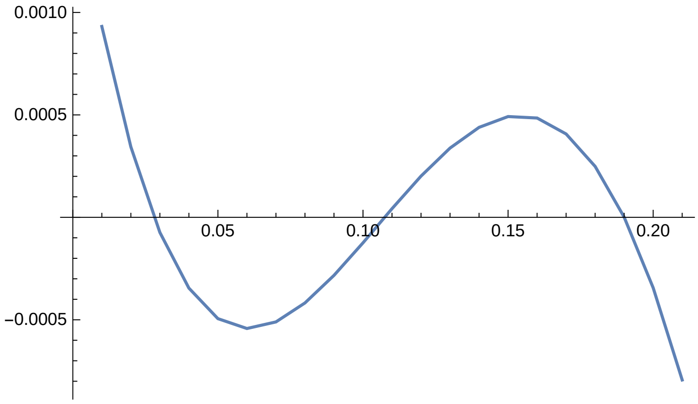
	DF	SS	MS	F-Statistic	P-Value
α	1	0.448441	0.448441	2.02659×10^{6}	6.37962×10^{-47}
α^2	1	0.000653945	0.000653945	2955.31	2.03123×10^{-21}
Error	18	3.98301×10^{-6}	2.21278×10^{-7}		
Total	20	0.449099			

In[53]:=
```
ans["RSquared"]
```

Out[53]= 0.999991

In[54]:=
```
ListPlot[{First /@ B, ans["FitResiduals"]} // Transpose, Joined → True]
```

Out[54]=



Cubic

In[72]:=
```
ans = LinearModelFit[B, {α, α^2, α^3}, α]
```

Out[72]= FittedModel[$-0.000184682 + 2.71403\,\alpha - 0.876013 \ll 1 \gg - 2.51929\,\alpha^3$]

In[56]:=
```
ans["ParameterTable"]
```

Out[56]=

	Estimate	Standard Error	t-Statistic	P-Value
1	−0.000184682	0.0000435935	−4.23645	0.000556129
α	2.71403	0.00167626	1619.09	1.51881×10^{-45}
α^2	−0.876013	0.0174903	−50.0857	6.61293×10^{-20}
α^3	−2.51929	0.0523377	−48.1353	1.29328×10^{-19}

In[57]:=
```
ans["ANOVATable"]
```

Out[57]=

	DF	SS	MS	F-Statistic	P-Value
α	1	0.448441	0.448441	2.62783×10^{8}	1.4879×10^{-62}
α^2	1	0.000653945	0.000653945	383 206.	1.90465×10^{-38}
α^3	1	3.954×10^{-6}	3.954×10^{-6}	2317.01	1.29328×10^{-19}
Error	17	2.90106×10^{-8}	1.70651×10^{-9}		
Total	20	0.449099			

In[58]:=
```
ans["RSquared"]
```

Out[58]= 1.

In[59]:=
```
InputForm[#] & /@ (ans["BestFitParameters"])
```

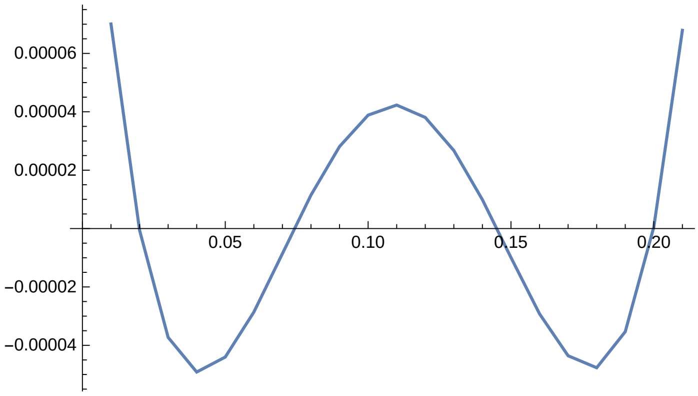
Out[59]=
```
{-0.00018468194429807232, 2.7140277221213913,
 -0.876012944101314, -2.51929256972748}
```

In[75]:=
```
ans[0.2]
```

Out[75]=
```
0.487426
```

In[60]:=
```
ListPlot[{First /@ B, ans["FitResiduals"]} // Transpose, Joined → True]
```

Out[60]=



Predict α given η, using cubic

In[61]:=
```
B2 = Reverse /@ B;
```

In[62]:=
```
B2
```

Out[62]=
```
{{0.0269355, 0.01}, {0.0537248, 0.02}, {0.0803424, 0.03},
 {0.106764, 0.04}, {0.132968, 0.05}, {0.158931, 0.06},
 {0.184632, 0.07}, {0.210053, 0.08}, {0.235174, 0.09},
 {0.259978, 0.1}, {0.284448, 0.11}, {0.308569, 0.12}, {0.332326, 0.13},
 {0.355706, 0.14}, {0.378697, 0.15}, {0.401286, 0.16}, {0.423462, 0.17},
 {0.445217, 0.18}, {0.466541, 0.19}, {0.487427, 0.2}, {0.507866, 0.21}}
```

In[63]:=
```
0.026935548879502003`
```

Out[63]=
```
0.0269355
```

In[64]:=
```
ans = LinearModelFit[B2, {η, η^2, η^3}, η]
```

Out[64]= FittedModel[$-0.000143935 + 0.37439\,\eta + \ll 21 \gg \eta^2 + 0.146135\,\eta^3$]

In[65]:=
```
ans["ParameterTable"]
```

Out[65]=

	Estimate	Standard Error	t-Statistic	P-Value
1	−0.000143935	0.0000419357	−3.43228	0.00317774
η	0.37439	0.000652649	573.647	6.95028×10^{-38}
η^2	0.00306868	0.00277202	1.10702	0.283705
η^3	0.146135	0.00339252	43.0756	8.42009×10^{-19}

In[66]:=
```
ans["ANOVATable"]
```

Out[66]=

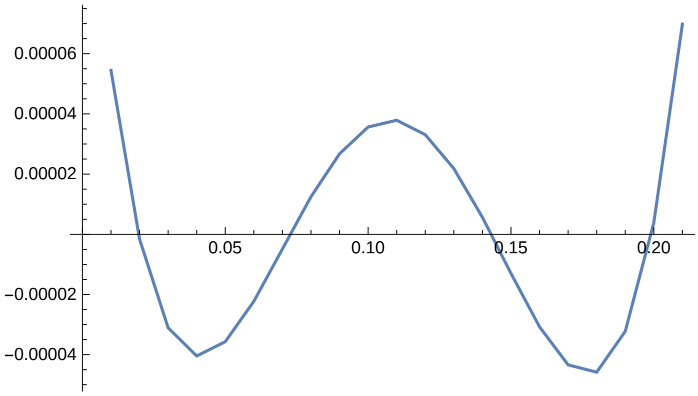
	DF	SS	MS	F-Statistic	P-Value
η	1	0.0768872	0.0768872	5.54301×10^{7}	8.26632×10^{-57}
η^2	1	0.000110207	0.000110207	79451.4	1.22331×10^{-32}
η^3	1	2.57378×10^{-6}	2.57378×10^{-6}	1855.51	8.42009×10^{-19}
Error	17	2.35807×10^{-8}	1.3871×10^{-9}		
Total	20	0.077			

In[67]:=
```
ans["RSquared"]
```

Out[67]= 1.

In[68]:=
```
ListPlot[{First /@ B, ans["FitResiduals"]} // Transpose, Joined → True]
```

Out[68]=



In[69]:=
```
InputForm[#] & /@ (ans["BestFitParameters"])
```

Out[69]=
```
{-0.00014393510482067046, 0.3743903928013429,
 0.003068677095495819, 0.14613484876904045}
```