Fuzzy Clustering

Theme

Initialization

In[111]:=
```
data = {{0, 0}, {1, 0}, {2, 1}, {2, 2}};
dataLabels = Subscript[bi["p"], #] & /@ Range[Length[data]];
clusterInit = {{1, 1}, {1, 2}};
bDefault = 2;
colorC1 = ■;
colorC2 = ■;
```

In[117]:=
```
cluster = clusterInit;
```

In[118]:=

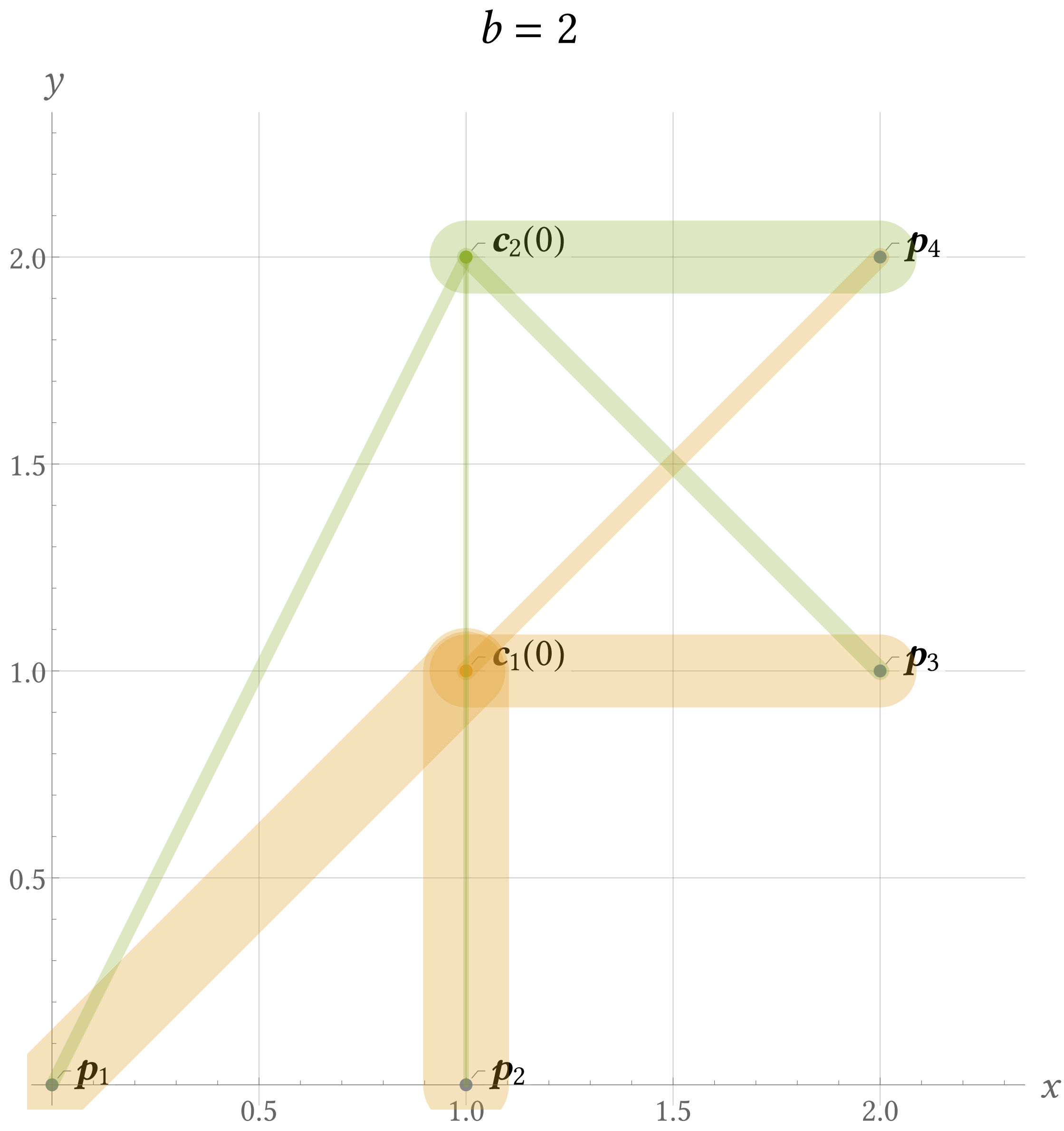
$$f_{b\_}[\mu\_, j\_] := \frac{1}{\sum_{i=1}^{\text{Length[cluster]}} \left(\frac{\text{Norm[data[[}\mu\text{]]-cluster[[j]]]}^2}{\text{Norm[data[[}\mu\text{]]-cluster[[i]]]}^2}\right)^{\frac{1}{b-1}}}$$

$$c_{j\_,b\_} := \frac{1}{\sum_{\mu=1}^{\text{Length[data]}} f_b[\mu, j]^b} \sum_{\mu=1}^{\text{Length[data]}} f_b[\mu, j]^b * \text{data}[[\mu]]$$

In[120]:=
```
calculateClusters[b_] := Module[{},
  (* It is important that both clusters are calculated before the definition is
   saved for the next iteration (otherwise, the second cluster would already use
     new definition for the first cluster when calculating the memberships) *)
  cluster = {c1,b, c2,b} // N (* Making the result numeric is important for performance *)
 ]
```

In[121]:=
```
plot[b_, t_] := ListPlot[{
   MapThread[Callout[#1, #2] &, {data, dataLabels}],
   {Callout[cluster[[1]], Row[{Subscript[bi["c"], 1], "(", t, ")"}]]},
   {Callout[cluster[[2]], Row[{Subscript[bi["c"], 2], "(", t, ")"}]]}
  },
  PlotTheme → "myTheme",
  GridLines → Automatic,
  AxesLabel → {x, y},
  AspectRatio → 1,
  PlotRange → {{-0.05, 2.35}, {-0.05, 2.35}},
  PlotLabel → Row[{Style["b", FontSlant → Italic], " = ", ToString[b]}],
  Epilog → {
    colorC1,
    Table[{CapForm["Round"], Opacity[0.3], AbsoluteThickness[50 * fb[j, 1]^b / (fb[j, 1]^b + fb[j, 2]^b)],
      Line[{cluster[[1]], data[[j]]}]}, {j, 1, Length[data]}],
    colorC2,
    Table[{CapForm["Round"], Opacity[0.3], AbsoluteThickness[50 * fb[j, 2]^b / (fb[j, 1]^b + fb[j, 2]^b)],
      Line[{cluster[[2]], data[[j]]}]}, {j, 1, Length[data]}]
   }
 ]
```

In[122]:=
```
plot[bDefault, 0]
```



Part 1: Re-Assignment Step

One can reduce the number of required calculations by making use of the fact that $f_{\mu,1} + f_{\mu,2} = 1$.

In[123]:=
```
Table[f_bDefault[μ, j], {μ, 1, Length[data]}, {j, 1, Length[cluster]}] // MatrixForm
```

Out[123]//MatrixForm=

$$\begin{pmatrix} \frac{5}{7} & \frac{2}{7} \\ \frac{4}{5} & \frac{1}{5} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

Part 2: Re-Calculation of the Cluster Centres

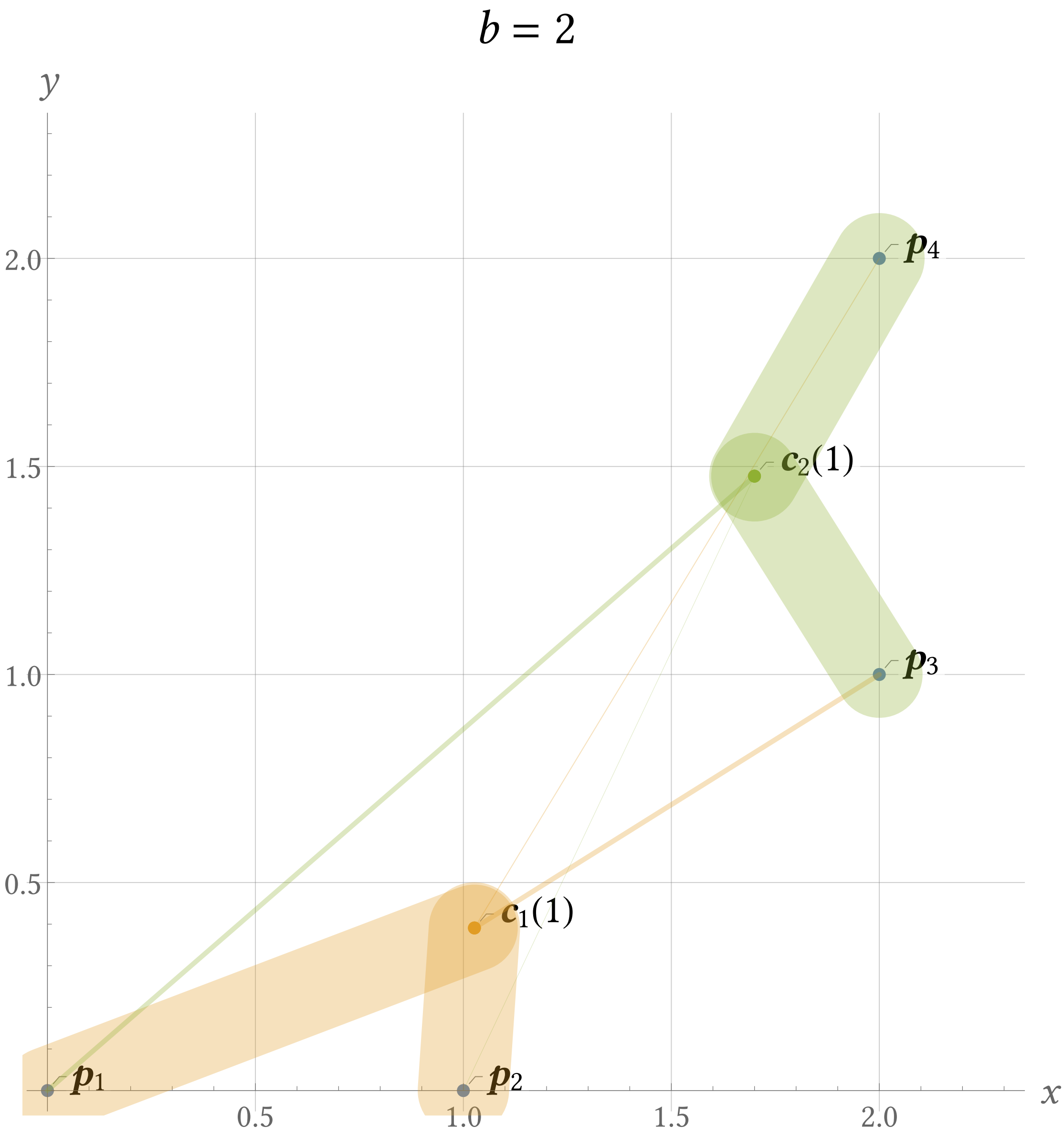
Cluster centres of iteration $t = 1$:

In[124]:=
```
calculateClusters[bDefault]
```

Out[124]=
```
{{1.02659, 0.390833}, {1.69984, 1.47669}}
```

In[125]:=
```
plot[bDefault, 1]
```

Out[125]=



Part 3: Cluster Result

Cluster centres of iteration $t = 100$:

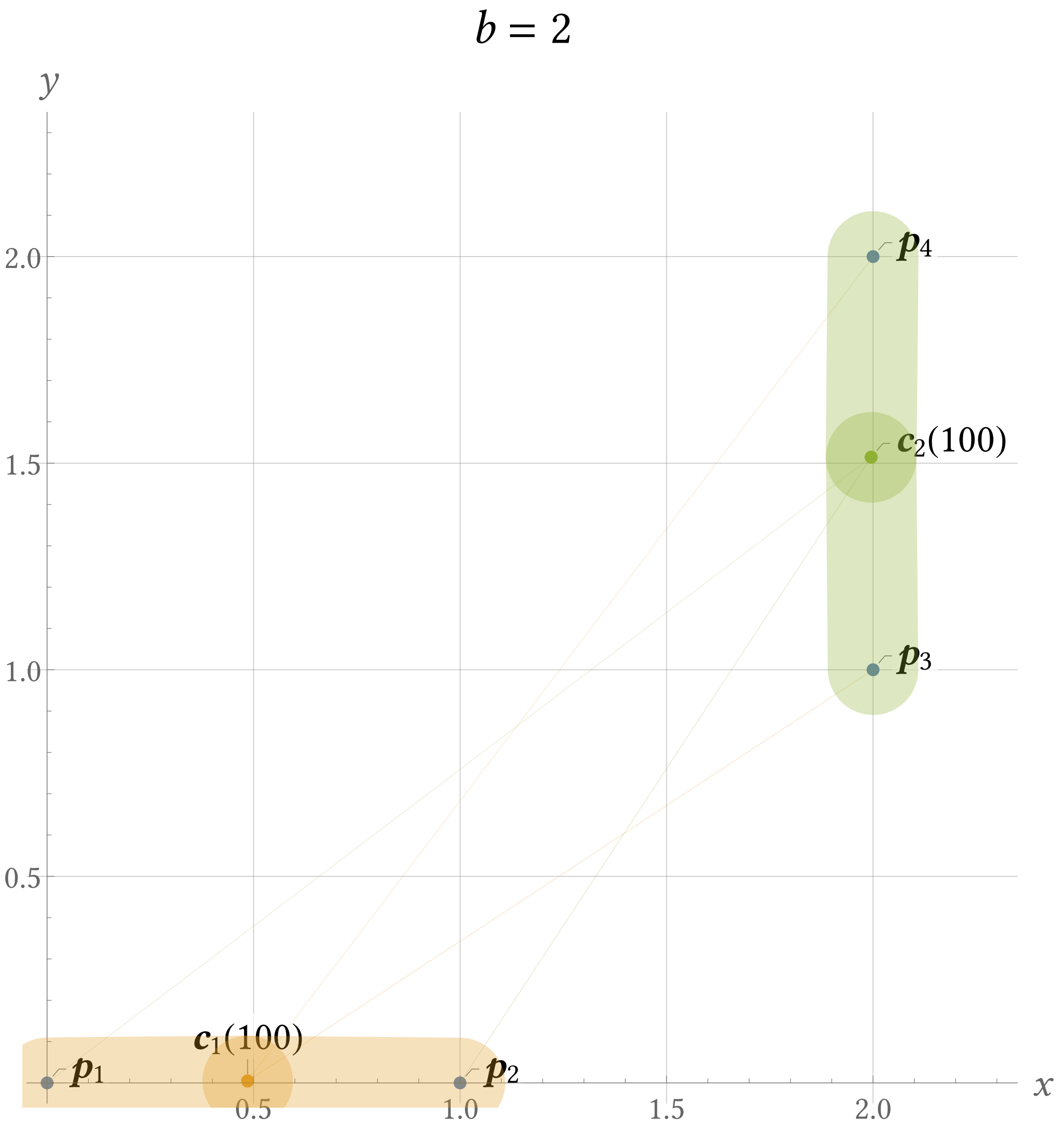
In[126]:=
```
Do[
 calculateClusters[bDefault];
 , {t, 2, 100}]
```

In[127]:=
```
cluster // N
```

Out[127]=
```
{{0.485528, 0.00456674}, {1.99543, 1.51447}}
```

The result is nearly the same as in crisp k-means clustering (where we end up with $\boldsymbol{c}_1 = (0.5, 0)$ and $\boldsymbol{c}_2 = (2, 1.5)$).

In[128]:=
```
plot[bDefault, 100]
```



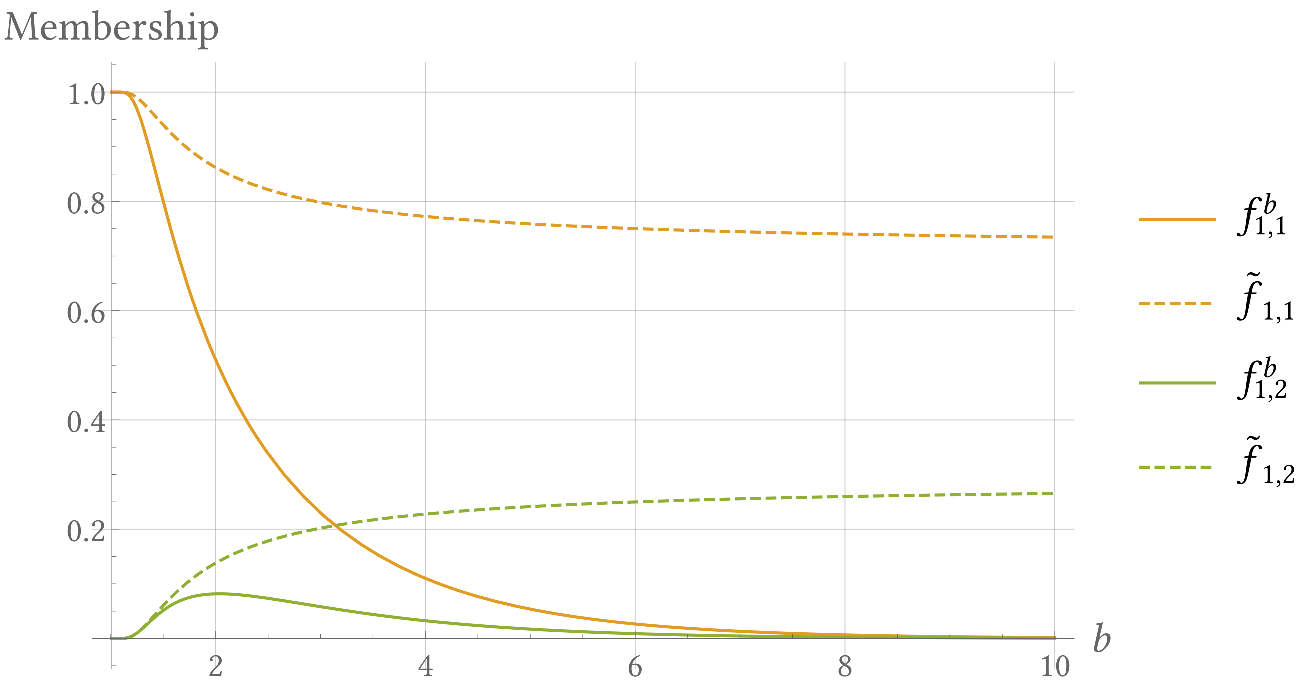
Part 4: Fuzzifier *b*

Graph Analysis

In[129]:=
```
cluster = clusterInit;
```

- Even though the values for $f_{1,j}$ tend to get smaller, we normalize them later with $\frac{1}{f_{1,1}^{b}+f_{1,2}^{b}}$ which does work as a counterforce to compensate for the small values.
- Regarding the overlap, it does indeed increase for larger values for *b* (the two dotted lines get closer to each other). This means that the membership is split more between the clusters.

In[130]:=
```
Plot[{f_b[1, 1]^b, f_b[1, 1]^b/(f_b[1, 1]^b + f_b[1, 2]^b), f_b[1, 2]^b, f_b[1, 2]^b/(f_b[1, 1]^b + f_b[1, 2]^b)}, {b, 1.01, 10},
 PlotTheme → "myTheme",
 GridLines → Automatic,
 PlotRange → Full,
 AxesLabel → {b, "Membership"},
 PlotLegends → {"f^b_{1,1}", "f̃_{1,1}", "f^b_{1,2}", "f̃_{1,2}"},
 PlotStyle → {colorC1, Directive[colorC1, Dashed], colorC2, Directive[colorC2, Dashed]}
]
```



The Limit $\lim_{b\to 1}$

The limits give us exact relationships meaning that we basically reduce to crisp *k*-means.

In[131]:=
```
Limit[f_b[1, 1]^b, b → 1]
```

Out[131]= 1

In[132]:=
```
Limit[f_b[1, 2]^b, b → 1]
```

Out[132]= 0

The critical part is the relationship (e.g. for $f_{1,1}$)

$$\frac{\|p_1 - c_1\|^2}{\|p_1 - c_2\|^2}$$

If the distance to the first cluster is lower than to the second cluster we get a value <1 and the exponent $\frac{1}{b-1}$ turns it into (basically) 0. Hence, we have $\frac{1}{1+0}=1$. On the other hand, if the distance to the first cluster is larger than to the second cluster, we have a value >1 and now the exponent turns it into a very large value resulting in $\lim_{a\to\infty}\frac{1}{1+a}=0$.

The Limit $\lim_{b\to\infty}$

In[143]:=
```
f11 = 1/(1 + (Norm[p1 - c1]^2/Norm[p1 - c2]^2)^(1/(b-1)));
f12 = 1/(1 + (Norm[p1 - c2]^2/Norm[p1 - c1]^2)^(1/(b-1)));
```

In[145]:=
```
f11Limit = Limit[f11^b/(f11^b + f12^b), b → ∞]
```

Out[145]= $\frac{\text{Norm}[-c_2+p_1]^2}{\text{Norm}[-c_1+p_1]^2+\text{Norm}[-c_2+p_1]^2}$

In[146]:=
```
f12Limit = Limit[f12^b/(f11^b + f12^b), b → ∞]
```

Out[146]= $\frac{\text{Norm}[-c_1+p_1]^2}{\text{Norm}[-c_1+p_1]^2+\text{Norm}[-c_2+p_1]^2}$

So, what does it mean? This basically tells us that in the other end of the extreme we take the relationship of the distances from the point to the cluster compared to the distances to all clusters. And as closer a point is to a cluster, as higher the membership is.

In[137]:=
```
f11Limit /. {x1 → data[[1]], c1 → cluster[[1]], c2 → cluster[[2]]}
```

Out[137]= $\frac{5}{7}$

In[138]:=
```
f12Limit /. {x1 → data[[1]], c1 → cluster[[1]], c2 → cluster[[2]]}
```

Out[138]= $\frac{2}{7}$

In[139]:=
```
Limit[fb[1, 1]^b/(fb[1, 1]^b + fb[1, 2]^b), b → ∞]
```

Out[139]= $\frac{5}{7}$

In[140]:=
```
Limit[fb[1, 2]^b/(fb[1, 1]^b + fb[1, 2]^b), b → ∞]
```

Out[140]= $\frac{2}{7}$

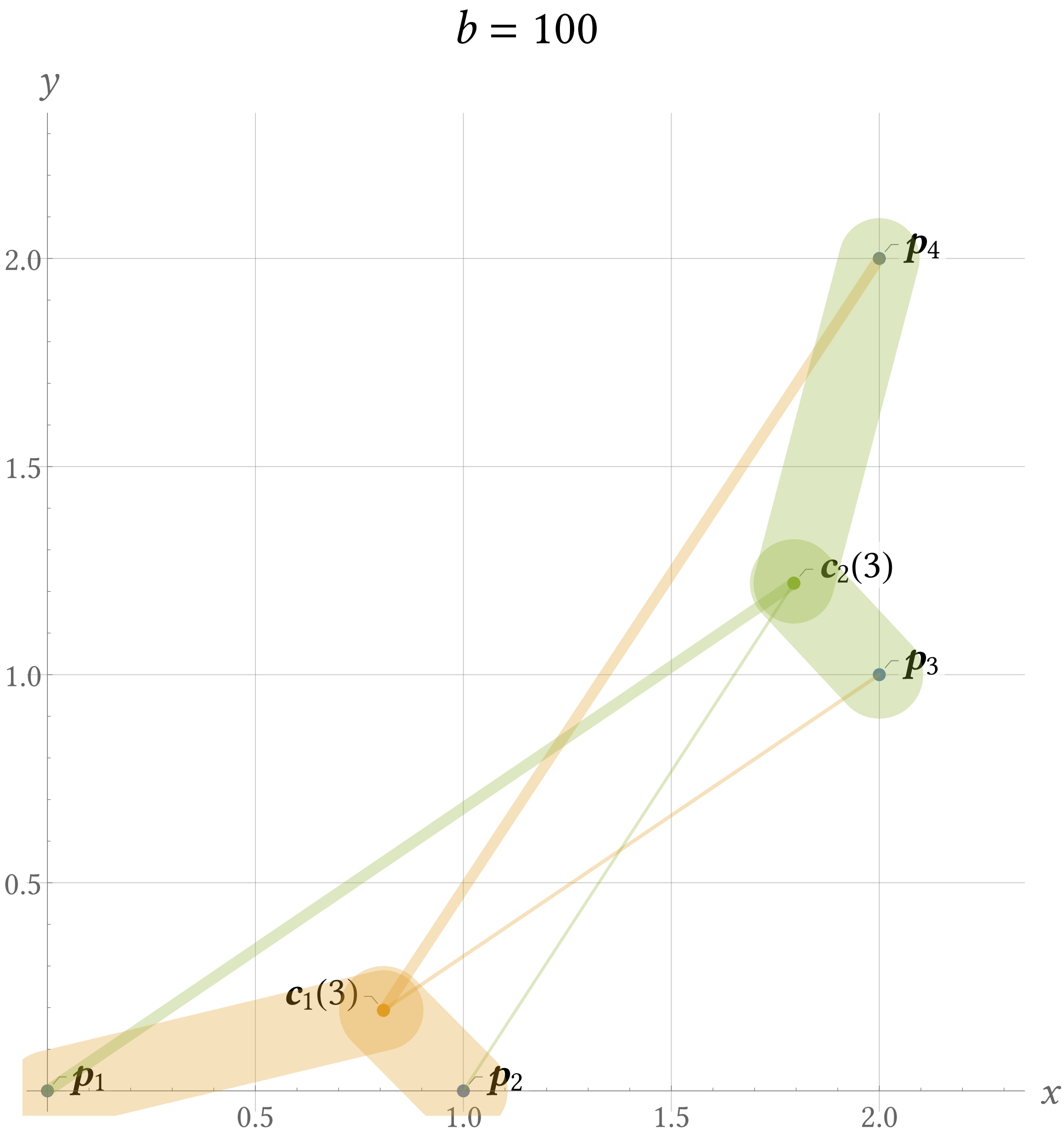
All together, the fuzzifier *b* controls the clustering between two extreme case. From the situation where we end up with crisp membership again to a situation where the result is solely based on the distances.

Higher Values for *b*

In the result for $b = 2$ we can see that the memberships are very crisp (two very thick and two very thin lines). If we now increase *b*, we amplify the overlap of the membership like shown in the previous analysis. This means that the two thin lines get bigger and pull the centroids more to themselves resulting in a situation as shown below.

In[147]:=
```
repeatClustering[100]
```

Out[147]=



In[141]:=
```
repeatClustering[b_] := Block[{iterations},
  cluster = {{1, 1}, {1, 2}};
  iterations = 3;

  Do[
   calculateClusters[b];
   , {i, 1, iterations}];

  plot[b, iterations]
 ]
```

In[142]:=
```
Manipulate[
 repeatClustering[bControl],
 {bControl, 1.1, 100}]
```

Out[142]=