

Creating some mock data

```
In[173]:= SeedRandom[1]
means = {{10, 20}, {30, 50}, {60, 70}};
kk = 3;
dd = 2;
nums = {150, 90, 200};
stds = {5, 8, 6};
rN[m_, s_] := Random[NormalDistribution[m, s]]
makePoint[mean_, std_] := {rN[mean[[1]], std], rN[mean[[2]], std]}

pts = Table[
  Table[makePoint[means[[k]], stds[[k]]], nums[[k]]],
  {k, 1, kk}];
```

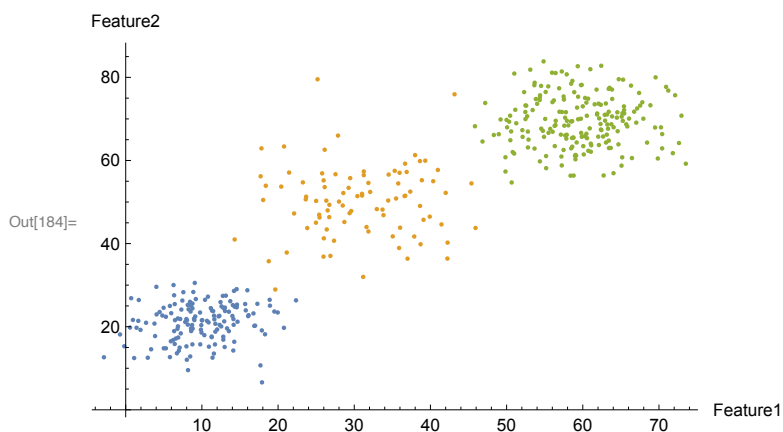
```
In[182]:= Length[pts]
```

```
Out[182]= 3
```

```
In[183]:= Length /@ pts
```

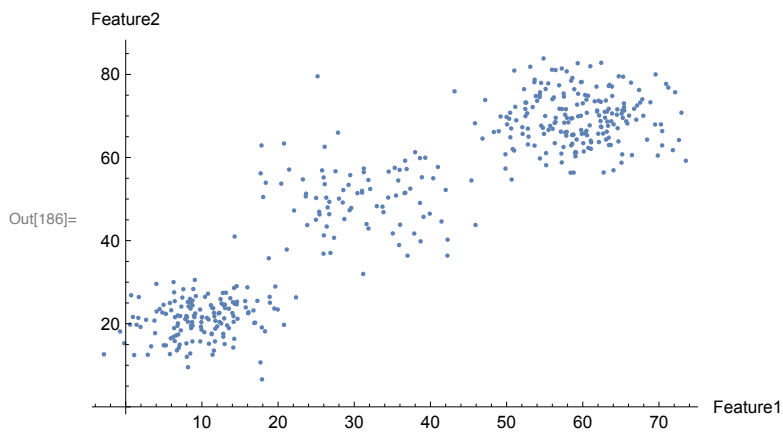
```
Out[183]= {150, 90, 200}
```

```
In[184]:= ListPlot[pts, AxesLabel → {"Feature1", "Feature2"}]
```



```
In[185]:= data = Flatten[Join[pts], 1];
```

```
In[186]:= ListPlot[data, AxesLabel → {"Feature1", "Feature2"}]
```



The k-means algorithm

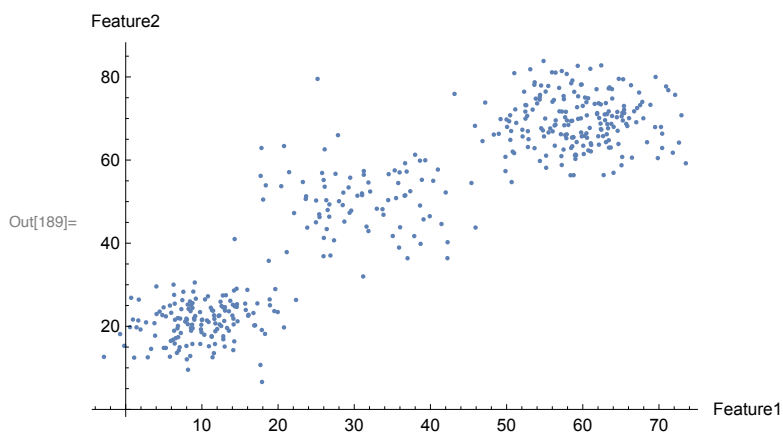
```
In[187]:= nn = Length[data]
```

Out[187]= 440

```
In[188]:= dd = Length[data[[1]]]
```

Out[188]= 2

```
In[189]:= ListPlot[data, AxesLabel → {"Feature1", "Feature2"}]
```



Let us assume 3 clusters

```
In[205]:= SeedRandom[0]
```

```
In[206]:= kk = 3
```

Out[206]= 3

```
In[207]:= minV = Min[data]
```

```
maxV = Max[data]
```

Out[207]= -2.85867

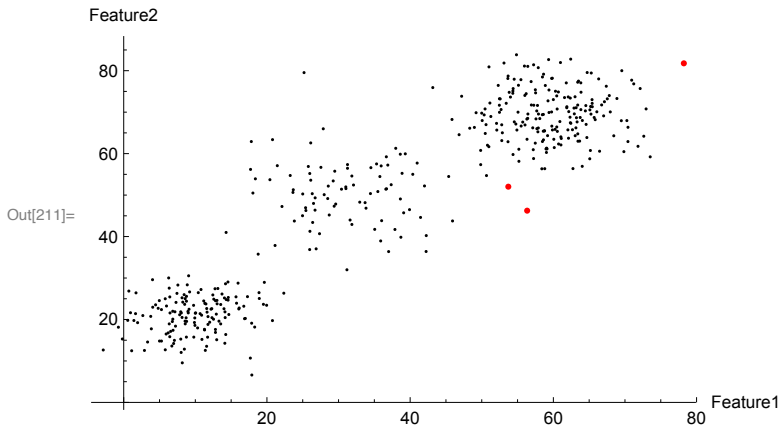
Out[208]= 83.8399

```
In[209]:= rP := RandomReal[{minV, maxV}]
```

```
In[210]:= centroids = {{rP, rP}, {rP, rP}, {rP, rP}}
```

```
Out[210]= {{53.7094, 52.0277}, {56.3403, 46.2432}, {78.2221, 81.7754}}
```

```
In[211]:= ListPlot[{data, centroids}, AxesLabel → {"Feature1", "Feature2"},  
PlotStyle → {{PointSize[0.005], Black}, {PointSize[0.01], Red}}]
```



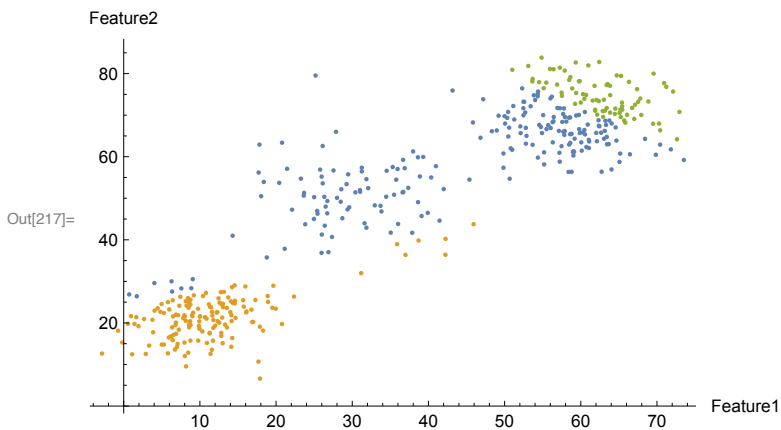
```
In[212]:= distsToCentroids[i_] := Table[Norm[data[[i]] - centroids[[k]]], {k, 1, kk}]
```

```
In[213]:= currentCluster[i_] := Ordering[distsToCentroids[i]][[1]]
```

```
clusterMapping = Table[currentCluster[i], {i, 1, nn}];
```

```
In[216]:= mappedPoints = Table[  
Pick[data, clusterMapping, k], {k, 1, kk}];
```

```
In[217]:= ListPlot[mappedPoints, AxesLabel → {"Feature1", "Feature2"}]
```



```
In[218]:= Length[mappedPoints]
```

```
Out[218]= 3
```

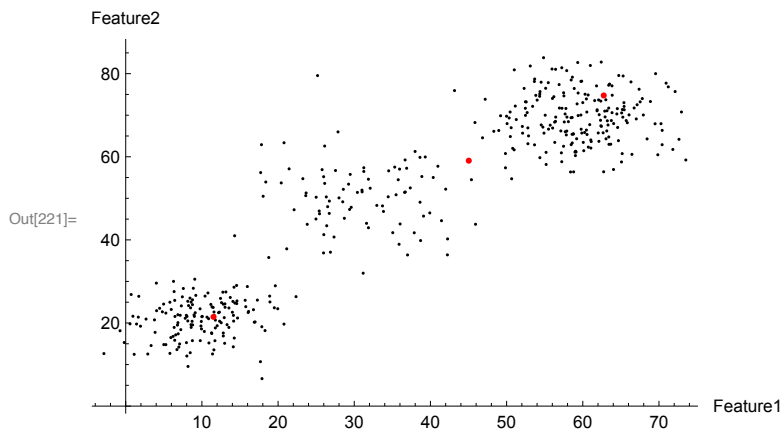
```
In[219]:= Length /@ mappedPoints
```

```
Out[219]= {213, 150, 77}
```

```
In[220]:= centroids = Table[Mean[mappedPoints[[k]]], {k, 1, kk}]
```

```
Out[220]= {{45.0347, 59.0577}, {11.5302, 21.4797}, {62.7491, 74.7532}}
```

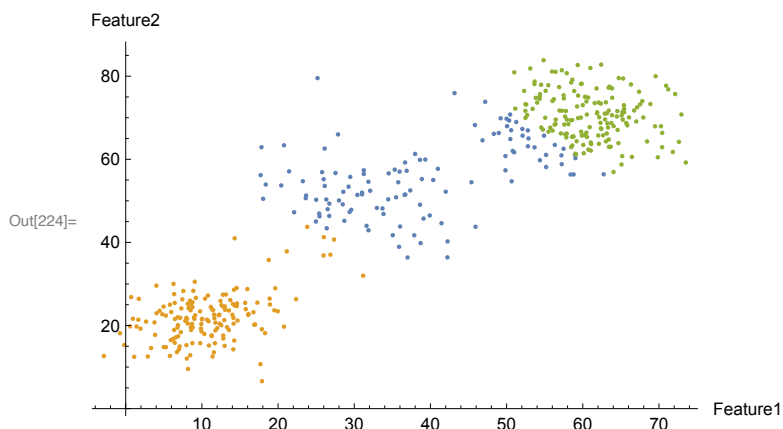
```
In[221]:= ListPlot[{data, centroids}, AxesLabel → {"Feature1", "Feature2"},
  PlotStyle → {{PointSize[0.005], Black}, {PointSize[0.01], Red}}]
```



```
In[222]:= clusterMapping = Table[currentCluster[i], {i, 1, nn}];
```

```
In[223]:= mappedPoints = Table[
  Pick[data, clusterMapping, k], {k, 1, kk}];
```

```
In[224]:= ListPlot[mappedPoints, AxesLabel → {"Feature1", "Feature2"}]
```



```
In[227]:= ? For
```

Out[227]=

Symbol
For[start, test, incr, body] executes start, then repeatedly evaluates body and incr until test fails to give True.

```
In[239]:= centroids = {{rP, rP}, {rP, rP}, {rP, rP}};
For[j = 1, j ≤ 10, j++,
  clusterMapping = Table[currentCluster[i], {i, 1, nn}];
  mappedPoints = Table[
    Pick[data, clusterMapping, k], {k, 1, kk}];
  centroids = Table[Mean[mappedPoints[[k]]], {k, 1, kk}];
  Print[Length /@ mappedPoints]
]
```

```
{218, 13, 209}  
{122, 99, 219}  
{81, 147, 212}  
{88, 151, 201}  
{87, 152, 201}  
{87, 152, 201}  
{87, 152, 201}  
{87, 152, 201}  
{87, 152, 201}  
{87, 152, 201}
```

```
In[241]:= ListPlot[mappedPoints, AxesLabel → {"Feature1", "Feature2"}]
```

