Functions

Naive Bayes Classification of Pets

```
In[156]:= petData = Import["/Users/CZMa/Downloads/Pet features (CSE514).csv"];
petData // Dimensions
```

```
Out[157]= {96, 6}
```

```
In[158]:= petData[[{1, 2, 3, -3, -2, -1}]] // MatrixForm
```

Out[158]//MatrixForm=

Timestamp	Color of the pet	Size of the pet	Is the
2022/09/26 12:06:23 PM EST	black and/or brown	Medium (5–20lb, or 2–10kg)	Warm-b
2022/09/26 2:08:47 PM EST	orange and/or red	Small (1–5lb, or 0.5–2kg)	Warm-b
2022/09/26 2:47:20 PM EST	white and/or gray	Medium (5–20lb, or 2–10kg)	Warm-b
2022/09/26 2:52:38 PM EST	orange and/or red	Tiny (<1lb or <0.5kg)	Warm-b
2022/09/26 2:56:07 PM EST	white and/or gray	Large (20–50lb, or 10–25kg)	Warm-b

Consolidating all "Other" values:

```
In[195]:= labelsOrig = {"Cat", "Dog", "Small mammal (hamster/gerbil/guinea pig/etc...)",
    "Reptile", "Fish or amphibian", "Bird", "Other"};
colorsOrig = {"black and/or brown", "orange and/or red",
    "mixed colors (with stripes/spots/patches)", "blond and/or yellow",
    "white and/or gray", "white and orange", "Other"};
```

```
In[197]:= petData[[2 ;;, -1]] =
    Map[If[Not[MemberQ[labelsOrig, #]], "Other", #] &, petData[[2 ;;, -1]]];
petData[[2 ;;, 2]] = Map[If[Not[MemberQ[colorsOrig, #]], "Other", #] &,
    petData[[2 ;;, 2]]];
```

P(class label)

```
In[188]:= labels = petData[[2 ;;, -1]] // DeleteDuplicates
For[i = 1, i ≤ Length[labels], i++,
 Print[labels[[i]], "\t",
  N[Length[Select[petData, #[[-1]] == labels[[i]] &]] / (Length[petData] - 1)]]
]
```

```
Out[188]= {Cat, Dog, Small mammal (hamster/gerbil/guinea pig/etc...),
    Reptile, Fish or amphibian, Bird, Other}
```

```
Cat     0.315789

Dog     0.463158

Small mammal (hamster/gerbil/guinea pig/etc...)     0.0210526

Reptile     0.0842105

Fish or amphibian     0.0526316

Bird     0.0315789

Other     0.0315789
```

Frequencies of color | class label

In[217]:=
```
colors = petData[[2 ;;, 2]] // DeleteDuplicates;
freqTable = Table[0, Length[colors], Length[labels]];
For[i = 1, i ≤ Length[labels], i++,
  For[j = 1, j ≤ Length[colors], j++,
   freqTable[[j, i]] =
     Length[Select[petData, #[[-1]] == labels[[i]] && #[[2]] == colors[[j]] &]];
  ]
 ];
labelMatrix[freqTable, labels, colors] // MatrixForm
```

Out[220]//MatrixForm=

	Cat	Dog	Small mammal (hamster/gerbil/gu
black and/or brown	9	8	0
orange and/or red	4	3	0
mixed colors (with stripes/spots/patches)	6	2	0
blond and/or yellow	2	12	1
white and/or gray	8	18	1
white and orange	0	1	0
Other	1	0	0

P(color | class label)

In[223]:=
```
freqTable = freqTable + 1;
labelMatrix[Map[N[#/Total[#]] &, freqTable // Transpose] // Transpose,
  labels, colors] // MatrixForm
```

Out[224]//MatrixForm=

	Cat	Dog	Small mammal (hamst
black and/or brown	0.25	0.172414	
orange and/or red	0.136364	0.0862069	
mixed colors (with stripes/spots/patches)	0.181818	0.0689655	
blond and/or yellow	0.0909091	0.241379	
white and/or gray	0.227273	0.344828	
white and orange	0.0454545	0.0517241	
Other	0.0681818	0.0344828	

Frequencies of sizes | class label

```
In[227]:= sizes = petData[2 ;;, 3] // DeleteDuplicates;
     freqTable = Table[0, Length[sizes], Length[labels]];
     For[i = 1, i ≤ Length[labels], i++,
       For[j = 1, j ≤ Length[sizes], j++,
        freqTable[j, i] = Length[Select[petData, #[-1] == labels[i] && #[3] == sizes[j] &]];
       ]
      ];
     labelMatrix[freqTable, labels, sizes] // MatrixForm
```

Out[230]//MatrixForm=

	Cat	Dog	Small mammal (hamster/gerbil/guinea pig/etc..
Medium (5–20lb, or 2–10kg)	19	19	0
Small (1–5lb, or 0.5–2kg)	8	12	1
Tiny (<1lb or <0.5kg)	2	1	1
Large (20–50lb, or 10–25kg)	0	9	0
Giant (>50lb, or >25kg)	1	3	0

P(sizes | class label)

```
In[231]:= freqTable = freqTable + 1;
     labelMatrix[Map[N[# / Total[#]] &, freqTable // Transpose] // Transpose,
       labels, sizes] // MatrixForm
```

Out[232]//MatrixForm=

	Cat	Dog	Small mammal (hamster/gerbil/guin
Medium (5–20lb, or 2–10kg)	0.571429	0.408163	0.142857
Small (1–5lb, or 0.5–2kg)	0.257143	0.265306	0.285714
Tiny (<1lb or <0.5kg)	0.0857143	0.0408163	0.285714
Large (20–50lb, or 10–25kg)	0.0285714	0.204082	0.142857
Giant (>50lb, or >25kg)	0.0571429	0.0816327	0.142857

Frequencies of temp | class label

```
In[233]:= temp = petData[2 ;;, 4] // DeleteDuplicates;
     freqTable = Table[0, Length[temp], Length[labels]];
     For[i = 1, i ≤ Length[labels], i++,
       For[j = 1, j ≤ Length[temp], j++,
        freqTable[j, i] = Length[Select[petData, #[-1] == labels[i] && #[4] == temp[j] &]];
       ]
      ];
     labelMatrix[freqTable, labels, temp] // MatrixForm
```

Out[236]//MatrixForm=

	Cat	Dog	Small mammal (hamster/gerbil/guinea pig/etc...)	Reptile
Warm-blooded	29	41	2	1
Cold-blooded	1	1	0	7
Other/Don't know	0	2	0	0

P(temp | class label)

In[237]:= `freqTable = freqTable + 1;`
`labelMatrix[Map[N[# / Total[#]] &, freqTable // Transpose] // Transpose,`
`  labels, temp] // MatrixForm`

Out[238]//MatrixForm=

	Cat	Dog	Small mammal (hamster/gerbil/guinea pig/etc.
Warm-blooded	0.909091	0.893617	0.6
Cold-blooded	0.0606061	0.0425532	0.2
Other/Don't know	0.030303	0.0638298	0.2

Frequencies of baths | class label

In[239]:= `baths = petData[[2 ;;, 5]] // DeleteDuplicates;`
`freqTable = Table[0, Length[baths], Length[labels]];`
`For[i = 1, i ≤ Length[labels], i++,`
`  For[j = 1, j ≤ Length[baths], j++,`
`   freqTable[[j, i]] = Length[Select[petData, #[[-1]] ⩵ labels[[i]] && #[[5]] ⩵ baths[[j]] &]];`
`  ]`
` ];`
`labelMatrix[freqTable, labels, baths] // MatrixForm`

Out[242]//MatrixForm=

	Cat	Dog	Small mammal (hamster/gerbil/guinea pig
Less frequently than once a year	5	1	0
More frequently than once a month	4	15	0
About once a month	13	27	1
About once a year	8	1	1

P(baths | class label)

In[243]:= `freqTable = freqTable + 1;`
`labelMatrix[Map[N[# / Total[#]] &, freqTable // Transpose] // Transpose,`
`  labels, baths] // MatrixForm`

Out[244]//MatrixForm=

	Cat	Dog	Small mammal (hamster/gerbil
Less frequently than once a year	0.176471	0.0416667	0.166667
More frequently than once a month	0.147059	0.333333	0.166667
About once a month	0.411765	0.583333	0.333333
About once a year	0.264706	0.0416667	0.333333