1. Presume we have learned a linear separator $\mathbf{w}$ and $b$ for logistic classification:

$$\mathbf{w} = \begin{bmatrix} -2 \\ 0 \\ 3 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad b = 5.$$ 

We also will use the sigmoid function.

(a) What class (0 or 1) is assigned to the following data points?

$$\mathbf{x}^1 = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 4 \\ -2 \end{bmatrix} \quad \Rightarrow \quad 1$$

$$\mathbf{x}^2 = \begin{bmatrix} 1 \\ 0 \\ -3 \\ 4 \\ 0 \end{bmatrix} \quad \Rightarrow \quad 0$$

(b) We wish to create a four-way classifier now – for classes 0, 1, 2, and 3. How can we use binary linear separators to distinguish between four classes?

One answer is:

Create a linear classifier to separate class 0 from not-0
Create a second linear classifier to separate class 1 from not-1
Create a third linear classifier to separate class 2 from not-2
Any data points in not-0, not-1, and not-2 will be auto-placed in class 3.
2. We wish to perform voice recognition – identifying who is currently speaking based on input audio information. One of three potential people is speaking – Jane, Maria, or Sam. We first classify speech using a maximum likelihood classifier using three features – pitch (sound frequency), speed (words-per-second), and volume. For each feature:
- the straight line indicates the distribution for Jane
- the small-dotted line indicates the distribution for Maria
- the big-dotted line indicates the distribution for Sam.

Presume the prior probabilities
\[ P(\text{speaker}=\text{Jane}) = 0.3, \quad P(\text{speaker}=\text{Maria})=0.5, \quad P(\text{speaker}=\text{Sam}) = 0.2 \]

We observe pitch=20, wordRate=0.5, volume = 6

What is the class, based on Posterior probability?

\[
\begin{align*}
\text{Jane: } & \quad 0.3 \times 0.7 \times 1 \times 0 = 0 \\
\text{Maria: } & \quad 0.5 \times 0.6 \times 0.03 \times 0.6 = 0.0054 \\
\text{Sam: } & \quad 0.2 \times 0 \times 2 \times 0.2 = 0
\end{align*}
\]

\textbf{Maria}
3. Presume we instead us another likelihood for duration \( d \) (words-per-sentence). It’s probability function follows the Rayleigh distribution:

\[
P_r(d|\sigma) = \frac{d}{\sigma} e^{-d^2/(2\sigma^2)}
\]

Example resulting shapes are shown to the right.

We wish to classify solely based on this feature. What is the maximum likelihood estimate of parameter \( \sigma \) for class speaker=Maria, to maximize:

\[
L = \prod_{i \in Maria} P_r(d_i|\sigma_{Maria})
\]

(Show your work for partial credit!)

\[
\prod_{i} \frac{d^i}{\sigma} e^{-\frac{d^i^2}{(2\sigma^2)}}
\]

\[
\sum_{i} \left( \log d^i - \log \sigma - \frac{d^i^2}{(2\sigma^2)} \right)
\]

Derivative:

\[
\sum_{i} \left( -\frac{1}{\sigma} + 2 \frac{d^i^2}{(2\sigma^3)} \right) = 0
\]

\[
\sum_{i} \left( \frac{d^i^2}{\sigma^3} \right) = \frac{N}{\sigma}
\]

\[
\sigma = \sqrt{\frac{\sum id^i^2}{N}}
\]
4. (20 points) Presume we have a np array variable \texttt{Data} with 5000 data points (as rows) and 40 features (as columns). (numpy: \texttt{Data.shape} is [5000, 40]). We also have a variable of class labels \texttt{Labels} with 5000 entries (as rows), each either 'A' or 'B', indicating the class for the data point in the corresponding row of \texttt{Data}. Finally, we have a function \texttt{Classify} that takes in a single 40-element row of features and outputs the predicted class label 'A' or 'B'. E.g.: \texttt{Classify(rowVector)} outputs an ‘A’ or a ‘B’

Write code to process the data points in \texttt{Data} and compute the percent correct labels output by the function \texttt{Classify}. Store the percent correct as a decimal between 0.0 and 1.0 in the variable \texttt{percentCorrect}.

One version of answer:

\begin{verbatim}
Count=0
for i in range(len(Data)):
    if Classify(Data[i,:])==Labels(i):
        Count=Count+1;

percentCorrect=Count/5000;
\end{verbatim}