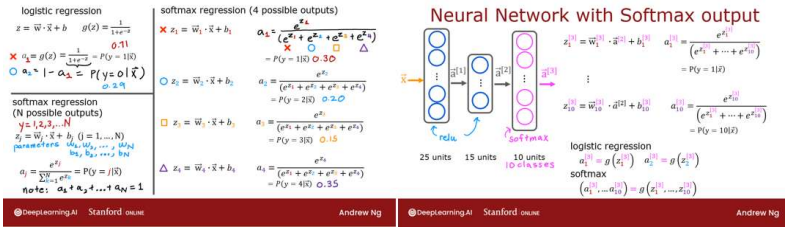
Optional Lab - Softmax Function

In this lab, we will explore the softmax function. This function is used in both Softmax Regression and in Neural Networks when solving Multiclass Classification problems.

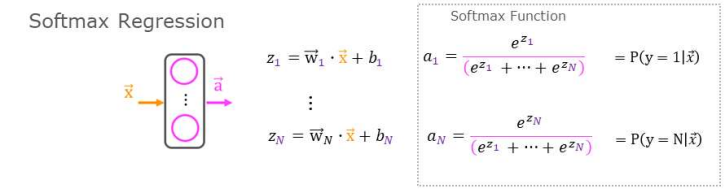


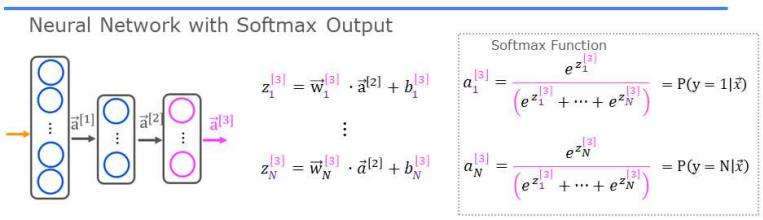
In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
plt.style.use('./deeplearning.mplstyle')
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from IPython.display import display, Markdown, Latex
from sklearn.datasets import make_blobs
%matplotlib widget
from matplotlib.widgets import Slider
from lab_utils_common import dlc
from lab_utils_softmax import plt_softmax
import logging
logging.getLogger("tensorflow").setLevel(logging.ERROR)
tf.autograph.set_verbosity(0)
```

Note: Normally, in this course, the notebooks use the convention of starting counts with 0 and ending with N-1, $\sum_{i=0}^{N-1}$, while lectures start with 1 and end with N, $\sum_{i=1}^{N}$. This is because code will typically start iteration with 0 while in lecture, counting 1 to N leads to cleaner, more succinct equations. This notebook has more equations than is typical for a lab and thus will break with the convention and will count 1 to N.

Softmax Function

In both softmax regression and neural networks with Softmax outputs, N outputs are generated and one output is selected as the predicted category. In both cases a vector $\mathbf{z}$ is generated by a linear function which is applied to a softmax function. The softmax function converts $\mathbf{z}$ into a probability distribution as described below. After applying softmax, each output will be between 0 and 1 and the outputs will add to 1, so that they can be interpreted as probabilities. The larger inputs will correspond to larger output probabilities.





The softmax function can be written:

$$a_j = \frac{e^{z_j}}{\sum_{k=1}^{N} e^{z_k}} \tag{1}$$

The output $\mathbf{a}$ is a vector of length N, so for softmax regression, you could also write:

$$\mathbf{a}(x) = \begin{bmatrix} P(y=1|\mathbf{x};\mathbf{w},b) \\ \vdots \\ P(y=N|\mathbf{x};\mathbf{w},b) \end{bmatrix} = \frac{1}{\sum_{k=1}^{N} e^{z_k}} \begin{bmatrix} e^{z_1} \\ \vdots \\ e^{z_N} \end{bmatrix} \tag{2}$$