## Model #2: Multinomial logistic regression 📈

recall $\sigma(t) = \frac{1}{1+e^{-t}} \cdot \frac{e^t}{e^t} = \frac{e^t}{e^t+1}$

- **Multinomial** logistic regression, also known as **softmax regression**, models the probability of belonging to **any class, given a feature vector $\vec{x}_i \in \mathbb{R}^{784}$.**

Think of it as a generalization of logistic regression.

Logistic regression: one parameter vector, $\vec{w}$

$$P(y_i = 1 \mid \vec{x}_i) = \sigma(\underbrace{\vec{w}}_{\text{param}} \cdot \underbrace{\text{Aug}(\vec{x}_i)}_{\text{augmented feature vec}}) = \frac{e^{\vec{w} \cdot \text{Aug}(\vec{x}_i)}}{e^{\vec{w} \cdot \text{Aug}(\vec{x}_i)} + 1}$$

Multinomial log reg: 10 parameter vectors, one per class, $\vec{w}_0, \vec{w}_1, \ldots, \vec{w}_9$!

$$\underbrace{P(y_i = 3}_{\text{prob that it's a 3}} \mid \underbrace{\vec{x}_i}_{\text{one image}}) = \frac{e^{\vec{w}_3 \cdot \text{Aug}(\vec{x}_i)}}{e^{\vec{w}_0 \cdot \text{Aug}(\vec{x}_i)} + e^{\vec{w}_1 \cdot \text{Aug}(\vec{x}_i)} + \ldots + e^{\vec{w}_9 \cdot \text{Aug}(\vec{x}_i)}}$$

denominator NORMALIZES so $\sum p_i = 1$

Let $p_j$ represent the modelled probability of class $j$, given a feature vector. Note that $j \in \{0, 1, \ldots, 9\}$.

Then, for instance:

$$p_0 = P(y_i = 0|\vec{x}_i) = \frac{e^{\vec{w}_0 \cdot \text{Aug}(\vec{x}_i)}}{\sum_{j=0}^{9} p_j}$$

wrong! should be

$$\sum_{j=0}^{9} e^{\vec{w}_j \cdot \text{Aug}(\vec{x}_i)}$$

?

- The **softmax** function is a generalization of the logistic function to multiple dimensions. Suppose $\vec{z} \in \mathbb{R}^d$. Then, the softmax of $\vec{z}$ is defined element-wise as follows:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{d} e^{z_j}}$$

[-5, 2, 4]

argmax = 2

[0, 0, 1]

biggest

[0.0001, 0.1192, 0.8807]

- For example, suppose $\vec{z} = \begin{bmatrix} -5 \\ 2 \\ 4 \end{bmatrix}$. Then:

$$\sigma(\vec{z}) = \begin{bmatrix} \sigma(\vec{z})_1 \\ \sigma(\vec{z})_2 \\ \sigma(\vec{z})_3 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{e^{-5}}{e^{-5}+e^2+e^4} \\ \frac{e^{2}}{e^{-5}+e^2+e^4} \\ \frac{e^{4}}{e^{-5}+e^2+e^4} \end{bmatrix}}_{\text{note the constant denominator!}} = \begin{bmatrix} 0.0001 \\ 0.1192 \\ 0.8807 \end{bmatrix}$$

largest prob, corresponds to the largest input