

Predicted_i = $y_{1,i} \cdot w_3 + y_{2,i} \cdot w_4 + b_3$
Now we can use chain rule to figure
out optimization

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

Where \vec{z} is an input vector and K is the number of
classes. All values in softmax will add to 1.

$$\arg \max |z| = \arg \max_i z_i$$