ICTS 4156: Introduction to ML
Razvan Bunescu
Lecture notes, April 21, 2021

1 Notes on Quiz

2 Notes on Homework

Use the formulas on slide 18 to compute $Z_2$, $A_2$, and $Z_3$ (the softmax logit scores). Use the code from the previous assignment to compute the softmax probabilities $A_3$ and the corresponding softmax loss. The only difference is that now softmax uses $A_2$ as input, instead of $X$.

3 Notes on NNs

- $a_3^{(2)} = f(W_{31}^{(1)} \ast x_1 + W_{32}^{(1)} \ast x_2 + W_{33}^{(1)} \ast x_3 + b_3^{(1)})$
- The input can be seen as layer $L_1$ composed of neuron activations. Let $a_1^{(1)} = x_1$. In general, $a_j^{(1)} = x_j$.
- $a_3^{(2)} = f(W_{31}^{(1)} \ast a_1^{(1)} + W_{32}^{(1)} \ast a_2^{(1)} + W_{33}^{(1)} \ast a_3^{(1)} + b_3^{(1)})$.
- $a_3^{(2)} = f(z_3^{(2)})$, where $z_3^{(2)} = W_{31}^{(1)} \ast a_1^{(1)} + W_{32}^{(1)} \ast a_2^{(1)} + W_{33}^{(1)} \ast a_3^{(1)} + b_3^{(1)}$.
- In general, $a_j^{(2)} = f(W_{j1}^{(1)} \ast a_1^{(1)} + W_{j2}^{(1)} \ast a_2^{(1)} + W_{j3}^{(1)} \ast a_3^{(1)} + b_j^{(1)})$.
- In general, $a_j^{(l+1)} = f(W_{j1}^{(l)} \ast a_1^{(l)} + W_{j2}^{(l)} \ast a_2^{(l)} + \ldots W_{js_l}^{(l)} \ast a_{s_l}^{(l)} + b_j^{(l)})$.
- In general, $a_j^{(l+1)} = f(z_j^{(l+1)})$ where $z_j^{(l+1)} = W_{j1}^{(l)} \ast a_1^{(l)} + W_{j2}^{(l)} \ast a_2^{(l)} + \ldots W_{js_l}^{(l)} \ast a_{s_l}^{(l)} + b_j^{(l)}$.
- Vectorized, $a_j^{(l+1)} = f(z_j^{(l+1)})$ where $z_j^{(l+1)} = W^{(l)}[j,:]^T a^{(l)} + b_j^{(l)}$.
- Fully vectorized, $a^{(l+1)} = f(z^{(l+1)})$ where $z^{(l+1)} = W^{(l)} a^{(l)} + b^{(l)}$.

$softmax(z) = [a_1, a_2, \ldots, a_K]$ where $a_j = \frac{exp(z_j)}{\sum_{k=1}^{K} exp(z_k)}$, assuming $z = [z_1, z_2, \ldots, z_K]$. 