

Mock Quiz 2

Nipun Batra and TA Sukruta, TA Rahul.

March 2024

1. Prove that sigmoid is a special case of softmax.
2. Prove that if we do not have non-linear activations (like sigmoid, ReLU), then the overall network no matter how deep is still equivalent to linear regression.
3. What is the domain and range of sigmoid? What is the domain range of tanh? Can you write tanh in terms of sigmoid?
4. We have an input image of size: $32 \times 32 \times 6$. We want to train for 100 class classification. Our MLP has 200 neurons in first layer, with ReLU activation and 120 neurons in the next layer before the output layer. What is the number of parameters in the model? Where might you encounter images containing 6 channels?
5. Derive the vectorised form of gradient for logistic regression cost function.
6. Why is the logistic regression cost function called cross-entropy? Contrast it with entropy we studied earlier in decision trees.
7. Why might you not prefer one-hot encoding for next-character prediction task?
8. Why do we not use the squared loss function for logistic regression?
9. Generalise the binary cross-entropy loss to the multi-class cross-entropy loss with K classes.
10. Show the formulae for focal loss and ask why it might be better suited for imbalanced datasets.