Other Extensions?

- Use a different loss function (why)?

    - Quadratic (Squared), Cross-entropy, Exponential, KL Divergence, etc.

- Use a different activation function (why)?

    - Sigmoid
    $$f(z) = \frac{1}{1 + exp(-z)}$$

    - Tanh
    $$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

    - Rectified Linear Unit (ReLU)
    $$f(z) = max(0, z)$$

References

References