1. A função softmax é definida pela equação abaixo:

,

Onde X é um vetor e n é a dimensão desse vetor. Ao aplicar a função softmax em um vetor, o resultado é um vetor normalizado em relação a si próprio, e que pode ser interpretado como probabilidade pois a soma do vetor resultante é 1 e quanto maior o valor no vetor original, maior o valor no vetor final.

1. A função ReLU é definida pela seguinte formulação:

Ou seja, assume valor 0 para x menor ou igual a 0, e assume o valor do próprio x caso contrário. Existem variações da ReLU, como a Leaky ReLU, que para valores iguais ou inferiores a 0 é definida como 0.01x e, como generalização, a Parametric ReLU, que é definida com ɑx.

1. As redes que se utilizam de convoluções apresentaram resultados melhores quando comparadas com as redes neurais que não as utilizam. O resultado obtido anteriormente foi uma acurácia de 97,71% no conjunto de teste.

Com dois modelos de redes neurais convolutivas, os resultados foram: 98.85 e 99.05. Resultados próximos dos listados na seção de redes convolutivas no site do dataset MNIST, onde o pior resultado é de 98.3%.