Gradient checking

<http://ufldl.stanford.edu/wiki/index.php/Gradient_checking_and_advanced_optimization>

Multivariable chain rule

<https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/differentiating-vector-valued-functions/a/multivariable-chain-rule-simple-version>

zeros vs zeros\_like

<https://stackoverflow.com/questions/27464039/why-the-performance-difference-between-numpy-zeros-and-numpy-zeros-like>