

1 Introduction

There are two main techniques for analyzing neural network prediction.

Sensitivity: It measures the local effect of the neurons for the given output. It is discrete and shows how much impact a neuron makes in deciding the output.

$$\Sigma_p(\frac{\delta f}{\delta x_p})^2 = ||\Delta_x f(x)||^2$$

Decomposition: It measures the global effect for the given output. It is continuous.

$$\Sigma_p[f(x)]_p = f(x)$$

In this blog, we will look into the decomposition analysis of neural networks.