1 Model

We begin with a general model that encapsulates many of the works we will discuss. For an input $\mathbf{x}$, we want to predict a response $\mathbf{y}^*$. Our prediction is a function f_θ of $\mathbf{x}$. The function f_θ is parameterized by θ. We will consider models of the form

$$f_\theta(\mathbf{x}) = y_\theta\left(\sum_i \alpha_\theta^{(i)} \phi_\theta^{(i)}(\mathbf{x})\right). \tag{1}$$

Here, $\mathbf{x}$ is an L-dimensional real vector and $\mathbf{y}^*$ is an M-dimensional real vector. The ϕ_i are basis functions that map from $\mathbb{R}^L$ to $\mathbb{R}^K$, and the α_i are real scalars. Additionally, the sum over i need not be finite. The coefficients α_i may depend on $\mathbf{x}$ and our parameters θ, but we suppress the former dependence for notational simplicity. The function y_θ maps from $\mathbb{R}^K$ to $\mathbb{R}^M$ and is parameterized by θ as well.

We will show how this model encapsulates the works we are interested in understanding.

2 Ingrosso et al. (2022)

The model in Ingrosso et al. (2022) sets $M = 1$, uses linear basis functions, and sets y_θ to be the mean function after applying a nonlinearity. That is,

$$\sum_i \alpha_\theta^{(i)} \phi_\theta^{(i)}(\mathbf{x}) = \Theta\mathbf{x} + b_\theta \tag{2}$$

$$y_\theta(\mathbf{x}) = \frac{1}{K}\mathbf{1}^\top \sigma(\mathbf{x}). \tag{3}$$

Here, Θ is a $K \times L$ matrix, b_θ is a K-dimensional vector, and σ is a nonlinear function applied elementwise.