Lecturer: Prof. Homrighausen Scribe: Ahmed Alawami

1 Nonparametric Regression

The goal: Suppose $Y \in \mathbb{R}$ and nonparametrically fit the regression function:

$$\mathbb{E}Y|X = f_*(X)$$

One approach is as follows:

1. Choose a 'suitable' basis $\{\phi_k\}_{k=1}^{\infty}$
2. Express $f_*(X) = \sum_{k=1}^{\infty} \beta_k \phi_k(X)$ (Note: $\beta_k = \langle f_*, \phi \rangle$)
3. Choose a 'suitable' K (smoothness control) and set $f_*^K(X) = \sum_{k=1}^{K} \beta_k \phi_k(X)$
4. Estimate β_k using least squares.

2 Nonparametric Regressions Problems

1. The basis is both fixed and independent of the data.
2. It doesn't work well for a large number of covariates.
3. For an 'insuitable' choice of basis, a large K is needed to adequetly represent the data.
4. It is not apparent data that needs two different basis to adequetly discribe it.

3 Basic Neural Network

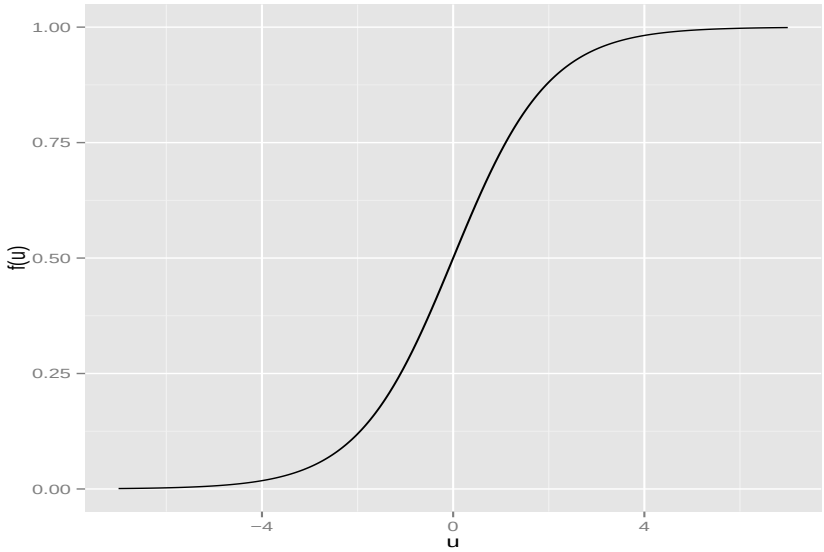
A Basic neural network representation is as follows:

$$L(\mu(X)) = \beta_0 + \sum_{k=1}^{K} \beta_k \sigma(\alpha_{k0} + \alpha_k^T X) \tag{1}$$

where L is a *link function* and $\mu(X) = \mathbb{E}Y|X$.

Note. For $L(\mu) = \mu$ the equation reduces to a regression in a transformed space. If we also had $\sigma(u) = u$, then it is reduced further to least squares regression.

Figure 1: Activation Function



4 Neural network Definitions

The main components of a neural network are the following:

1. The *derived features* (or *hidden units*) $Z_k = \sigma(\alpha_{k0} + \alpha_k^T X)$.
2. The *activation function* $\sigma(u)$.
3. The number of hidden units K (which is a tuning parameter).

5 Activation Function

The most common activation function is the *sigmoid function*.

$$\sigma(u) = (1 + \exp(-u))^{-1} \tag{2}$$

Note. The activation function generates a feature map.

6 Hierarchical Model

We can express neural networks as a hierarchical model as follows:

$$
\begin{aligned}
Z_k &= \sigma(\alpha_{k0}); && k = 1, \cdots, K \\
W_g &= \beta_{g0} + \beta_g^T Z; && g = 1, \cdots, G \\
\mu_g(X) &= L^{-1}(W_g)
\end{aligned}
$$

Note. For $L(u) = u$ and $G = 1$ we just have regression. We can do classification with G classes by choose L to be the logit function.