5 Fully Connected Neural Networks

84 Example (Concrete Strength)

(a) Use linear regression to predict the compressive strength of concrete using the features listed in the file `Concrete_train`[8]. Compare the speeds of the Adam optimizer to stochasitic gradient descent (SGD). Would you characterize linear regression as under or over-fitting the data.

(b) Add one or more 8 node (output) fully-connected layers and repeat part (a).

(c) Add one or more 8 node (output) fully-connected layers separated by ReLU activation layers and repeat part (a).

Since the composition of linear functions is linear, not much is gained by stacking linear layers.

85 Example (Stacking Linear Layers)

$$\mathbf{X} \to \boxed{\mathbf{W}_0} \to \mathbf{Z}_1 \to \boxed{\mathbf{w}_1} \to \hat{\mathbf{y}}$$

$$\begin{aligned} \mathbf{Z}_1 &= \mathbf{X}\mathbf{W}_0 \\ \hat{\mathbf{y}} &= \mathbf{Z}_1\mathbf{w}_1 \end{aligned}$$

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{X}\mathbf{W}_0\mathbf{w}_1 \\ &= \mathbf{X}\mathbf{w} \end{aligned}$$

where $\mathbf{w} = \mathbf{W}_0\mathbf{w}_1$. Thus, two stacked linear layers can be represented by a single linear layer.

---

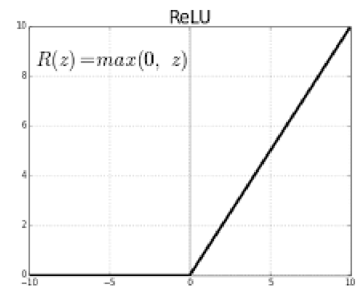
[8]Source: Prof. I-Cheng Yeh, Department of Information Management Chung-Hua University, Hsin Chu, Taiwan 30067, R.O.C., e-mail:icyeh@chu.edu.tw, TEL:886-3-518651y1

To increase the learning capacity of a network, linear layers must be separated by nonlinear layers. One of the simplest possible nonlinear layers is the **ReLU layer**.

86 Definition (ReLU)

Rectified Linear Units (ReLU's) are among the simplest of nonlinear functions:

$$\text{ReLU}(x) = \max\{0, x\}$$



---

87 Example (ReLU)

Assume $Y = \text{ReLU}(X)$ where X is the input data shown below. Determine Y.

x_1	x_2	x_3	y_1	y_2	y_3
3.1	−1.2	0.1			
−4.4	−0.2	0.4			

---

88 Definition (Fully-Connected (Dense) Regression Networks)

A **fully-connected (dense) regression network** is a neural network consisting of L linear layers separated by nonlinear layers such as the ReLU layer. Each linear layer has N_k, $k = 0, 1, \ldots, L-1$, **nodes** (also called **outputs**).

$$\underset{n\times(d+1)}{\mathbf{X}} \to \boxed{\underset{(d+1)\times N_0}{\mathbf{W}_0}} \to \underset{n\times(N_0+1)}{\mathbf{Z}_1} \to \boxed{f_{\text{ReLU}}} \to \underset{n\times(N_0+1)}{\mathbf{A}_1} \to$$

$$\boxed{\underset{(N_0+1)\times N_1}{\mathbf{W}_1}} \to \underset{n\times(N_1+1)}{\mathbf{Z}_2} \to \boxed{f_{\text{ReLU}}} \to \underset{n\times(N_1+1)}{\mathbf{A}_2} \to \cdots \to$$

$$\underset{n\times(N_{L-2}+1)}{\mathbf{A}_{L-1}} \to \boxed{\underset{(N_{L-2}+1)\times N_{L-1}}{\mathbf{W}_{L-1}}} \to \underset{n\times N_{L-1}}{\hat{\mathbf{Y}}}$$