Data (Input) · **First Layer (3 units)** · **Activation Function** · **Second Layer (2 units)** · **Output**

x_1 x_2 x_3 x_4

First unit
$$v_{i,1}^{(1)} = x_i' w_1^{(1)} + b_1^{(1)}$$

Second unit
$$v_{i,2}^{(1)} = x_i' w_2^{(1)} + b_2^{(1)}$$

Third unit
$$v_{i,3}^{(1)} = x_i' w_3^{(1)} + b_3^{(1)}$$

Nonlinear Transform

Output
$$o_{i,k}^{(1)} = f^{(1)}(v_{i,k})$$

Yields a vector $o_i^{(1)}$ for each data point i

output sent to next layer

First unit
$$v_{i,1}^{(2)} = o_i^{(1)'} w_1^{(2)} + b_1^{(2)}$$

Second unit
$$v_{i,2}^{(2)} = o_i^{(1)'} w_2^{(2)} + b_2^{(2)}$$

Aggregation via linear mapping with possible posterior activation

Yields one point for each initial occurrence

Notation:

Weights
$$w_k^{(l)}$$
index of layer
index of unit

Intermed. Values
$$v_{i,k}^{(l)}$$
index of layer
index of unit
index of occurrence

Activation functions are usually indexed according to the layer they succeed