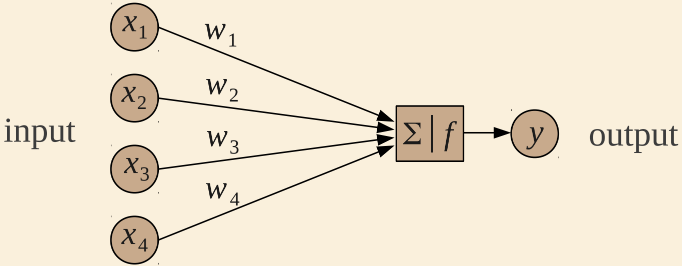
Lecture 6

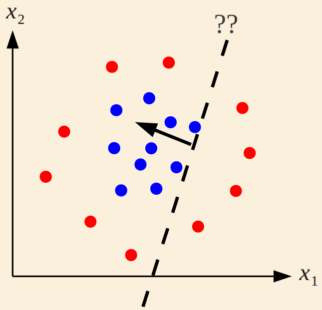
Stephanie Brandl

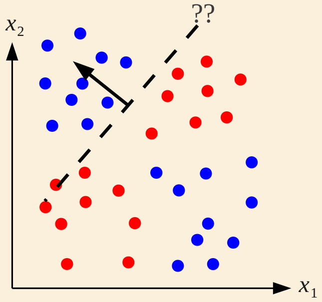
One-layer networks



$$y = f(x_1 w_1 + x_2 w_2 + x_3 w_3 + x_4 w_4 + b) = f(\mathbf{w} \cdot \mathbf{x} + b)$$

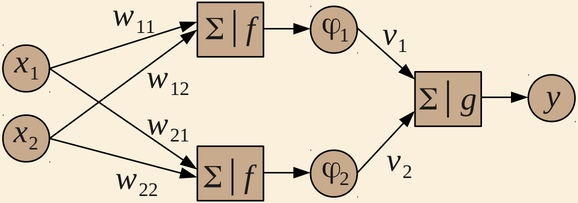
Limitation of one-layer networks





One-layer networks are **not powerful enough** to solve such problems. We need more sophisticated models.

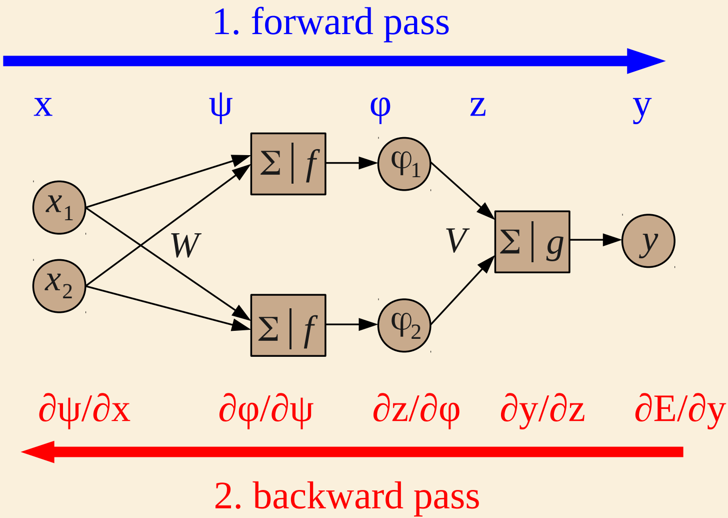
Multilayer networks



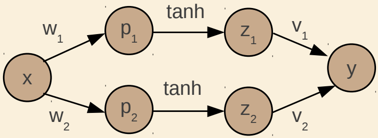
$$\varphi_1 = f\left(x_1 w_{11} + x_2 w_{12} + b_1\right)$$
$$\varphi_2 = f\left(x_1 w_{21} + x_2 w_{22} + b_2\right)$$
$$y = g\left(\varphi_1 v_1 + \varphi_2 v_2 + c\right)$$

Matrix form:
$$y = g\left(V \cdot f\left(W \cdot x\right)\right)$$

Learning with a multilayer network



Learning with a multilayer network



$$y = v_1 \tanh(w_1 x) + v_2 \tanh(w_2 x)$$

$$\text{error:} \quad E = \frac{1}{2}(y-t)^2$$

$$\tanh'(x) = 1 - \tanh(x)^2$$

$$\begin{aligned}
\frac{\partial E}{\partial w_1} &= \frac{\partial}{\partial w_1} \frac{1}{2}(y-t)^2 \\
&= \frac{\partial\, 1/2(y-t)^2}{\partial y} \cdot y'(w_1) \\
&= (y-t) \cdot \left[\frac{\partial y}{\partial z_1} \cdot z_1'(w_1) + \frac{\partial y}{\partial z_2} \cdot z_2'(w_1)\right] \\
&= (y-t) \cdot \left[v_1 \frac{\partial z_1}{\partial p_1} p_1'(w_1) + v_2 \frac{\partial z_2}{\partial p_2} p_2'(w_1)\right] \\
&= (y-t)\left(v_1(1-z_1^2)x + v_2(1-z_2^2)0\right)
\end{aligned}$$

Neural Networks are powerful algorithms which are used to solve many different machine learning problems. How fast and how reliable you reach your solution depends on different parameters such as:

i) initializing weights
   e.g. $w \sim \mathcal{N}(0, 1/\sqrt{n})$, s.t. $-1/\sqrt{n} < w < 1/\sqrt{n}$
ii) activation functions
   e.g. sigmoid function, linear nodes, soft-max
iii) number of hidden layers and nodes
   e.g. 22 in GoogleNet (100 000 parameters)
iv) learning rate η