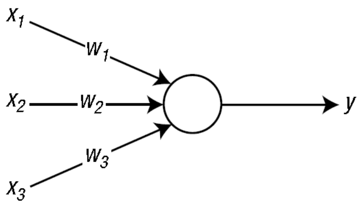
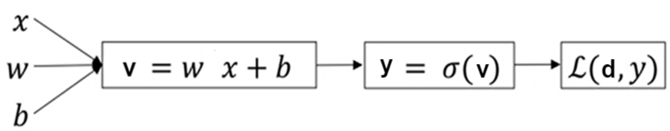
Forward and backward propogation

Example 1: Logistic regression



Computation graph



$v = wx + b$
$y = \sigma(v)$
$\mathscr{L}(y, d) = -d \log y - (1 - d) \log(1 - y)$

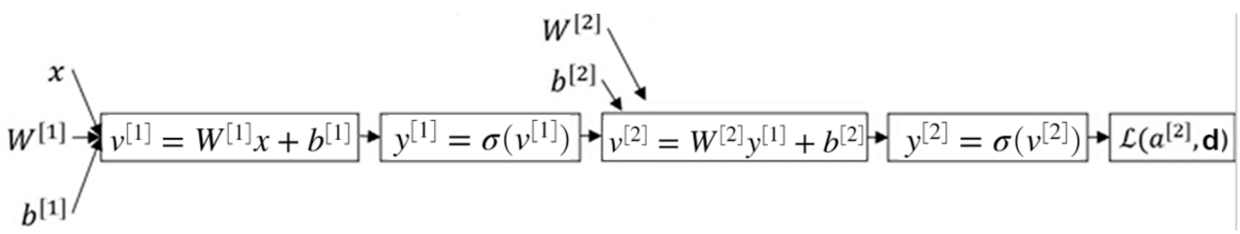
$$dy = \frac{d\mathscr{L}(y, d)}{dy} = -\frac{d}{y} + \frac{1 - d}{1 - y} = \frac{y - d}{y(1 - y)}$$

$$dv = \frac{\partial \mathscr{L}(y, d)}{\partial v} = \frac{\partial \mathscr{L}(d, y)}{\partial y}\frac{\partial y}{\partial v} = dy \cdot \sigma'(v) = dy \cdot y(1 - y) = y - d$$

$$dw = \frac{\partial \mathscr{L}}{\partial v}\frac{\partial v}{\partial w} = dv x^T$$

$$db = \frac{\partial \mathscr{L}}{\partial v}\frac{\partial v}{\partial b} = dv$$

Example 2



$v^{[1]} = W^{[1]}x + b^{[1]}$

$y^{[1]} = \sigma(v^{[1]})$

$v^{[2]} = W^{[2]}y^{[1]} + b^{[2]}$

$y^{[2]} = \sigma(v^{[2]})$

$\mathscr{L}(y^{[2]}, d) = -d \log y^{[2]} - (1-d)\log(1-y^{[2]})$

$dy^{[2]} = \dfrac{y^{[2]} - d}{y^{[2]}(1-y^{[2]})}$

$dv^{[2]} = y^{[2]} - d$ (对应code $e = d - y, \delta = e$)

$dW^{[2]} = dz^{[2]} y^{[1]T} \left(dW_{ij}^{[2]} = dv_i^{[2]} \dfrac{\partial v_i^{[2]}}{\partial W_{ij}^{[2]}} = dv_i^{[2]} y_j^{[1]} \right)$

$db^{[2]} = dv^{[2]}$

$dy^{[1]} = W^{[2]T} dv^{[2]} \left(dy_i^{[1]} = \sum_j dv_j^{[2]} \dfrac{\partial v_j^{[2]}}{\partial y_i^{[1]}} = \sum_j dv_j^{[2]} W_{ji} \right)$ (对应e1)

$dv^{[1]} = dy^{[1]} \cdot \sigma'(v^{[1]})$ (elementwise product,对应$\delta 1$)

$dW^{[1]} = dv^{[1]} x^T$

$db^{[1]} = dv^{[1]}$

Check the dimensions

input layer: $n^{[0]}$ nodes, hidden layer: $n^{[1]}$ nodes, output layer: $n^{[2]}$ nodes

$x - n^{[0]} \times 1; W^{[1]} - n^{[1]} \times n^{[0]}; b^{[1]} - n^{[1]} \times 1; W^{[2]} - n^{[2]} \times n^{[1]}; b^{[2]} - n^{[2]} \times 1$

向量对向量求导: $\dfrac{\partial Ax}{\partial x} = A^T$

$$A_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \dots a_{1n} \\ a_{21} & a_{22} & \dots a_{2n} \\ \dots & \dots & \ddots \quad \dots \\ a_{m1} & a_{m2} & \dots a_{mn} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}, \text{那么} Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{bmatrix}_{m\times 1}$$

$$\frac{\partial Ax}{\partial x} = \begin{bmatrix} a_{11} & a_{21} & \dots a_{m1} \\ a_{12} & a_{22} & \dots a_{m2} \\ \dots \\ a_{1n} & a_{2n} & \dots a_{mn} \end{bmatrix} = A^T$$

对谁求导数，就以谁（分母）作为主序，得出结果。比如这里x是列向量，求Ax关于x的导数，那么对x的每个分量分别求偏导数，排成一列（同x一样是列向量）

矩阵化形式（batch or minibatch method, m samples）

$X = [x^{(1)} \ \ x^{(2)} \cdots x^{(m)}]$-- $n^{[0]} \times m$ matrix, i is the index of sample

Forward

$$V^{[1]} = W^{[1]}X + b^{[1]}$$
$$Y^{[1]} = \sigma(V^{[1]})$$
$$V^{[2]} = W^{[2]}Y^{[1]} + b^{[2]}$$
$$Y^{[2]} = \sigma(V^{[2]})$$
$$\mathscr{L}(Y^{[2]}, D) = \frac{1}{m}\sum_{i=1}^{m} \mathscr{L}(Y^{[2](i)}, D^{(i)})$$

Backpropagation

$$dV^{[2]} = Y^{[2]} - D$$
$$dW^{[2]} = \frac{1}{m} dV^{[2]} Y^{[1]^T}$$
$$db^{[2]} = \frac{1}{m}\sum_{i=1}^{m} dV^{[2](i)}$$
$$dV^{[1]} = W^{[2]^T} dY^{[2]} \cdot \sigma'(Y^{[1]})$$
$$dW^{[1]} = \frac{1}{m} dV^{[1]} X^T$$
$$db^{[1]} = \frac{1}{m}\sum_{i=1}^{m} dV^{[1](i)}$$