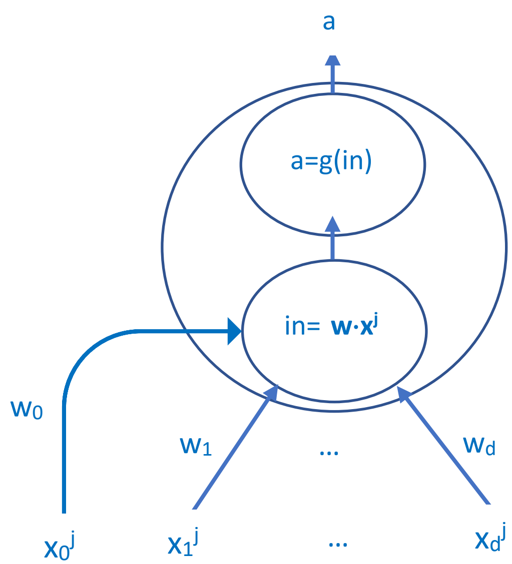

Perceptron:

$$g(z) = \text{if } z \geq 0 \text{ then } 1 \text{ else } 0$$

Logistic Regression:

$$g(z) = \frac{1}{1+e^{-z}}$$

Perceptron Learning Algorithm

Initialize $w_0,\dots,w_d$
Until <stopping condition>
    Randomly order the data $<x^1, \dots, x^N>$
        For j = 1 to N do
            if $\mathbf{w}\cdot\mathbf{x}^j \geq 0$ then h=1 else h=0
            For i = 0 to d do
                $w_i = w_i + \alpha \times (y^j - h) \times x^j_i$

Derived by Rosenblatt based on McCulloch/Pitts model of neurons, Hebbian model of learning
(not covered in class)

Logistic Regression Learning Algorithm

Initialize $w_0,\dots,w_d$
Until <stopping condition>
    Randomly order the data $<x^1, \dots, x^N>$
        For j = 1 to N do
            $h = \frac{1}{1+e^{-\mathbf{w}\cdot\mathbf{x}^{\mathbf{j}}}}$
            For i = 0 to d do
                $w_i = w_i + \alpha \times (y^j - h) \times h \times (1 - h) \times x^j_i$

Derived formula for gradient descent – minimizing

$$Loss_D(\mathrm{h}_{\mathbf{w}}) = \sum_{\mathrm{j}=1}^{\mathrm{N}} \left(\mathrm{y}^{\mathrm{j}} - \mathrm{h}_{\mathbf{w}}(\mathrm{x}^{\mathrm{j}})\right)^2$$

and then doing updates example by example
(*Stochastic Gradient Descent)*

Gradient Descent Learning Algorithm
(Not really used)

Initialize $w_0,\dots,w_d$
Until <stopping condition>
  Randomly order the data $\langle x^1, \dots, x^N \rangle$
    $\langle temp_0, \dots, temp_d \rangle = \langle w_0, \dots, w_d \rangle$
    For i = 0 to d do
      $h = \frac{1}{1+e^{-\mathbf{w}\cdot\mathbf{x}^{\mathbf{j}}}}$
      For j = 1 to N do
        $temp_i = temp_i$
        $+ \alpha \times (y^j - h) \times h \times (1 - h) \times x^j_i$
    $\langle w_0, \dots, w_d \rangle = \langle temp_0, \dots, temp_d \rangle$

Logistic Regression Learning Algorithm
(Stochastic Gradient Descent)

Initialize $w_0,\dots,w_d$
Until <stopping condition>
  Randomly order the data $\langle x^1, \dots, x^N \rangle$
    For j = 1 to N do
      $h = \frac{1}{1+e^{-\mathbf{w}\cdot\mathbf{x}^{\mathbf{j}}}}$
      For i = 0 to d do
        $w_i = w_i$
        $+ \alpha \times (y^j - h) \times h \times (1 - h) \times x^j_i$