# Representing Random Forests as Artificial Neural Networks

# History

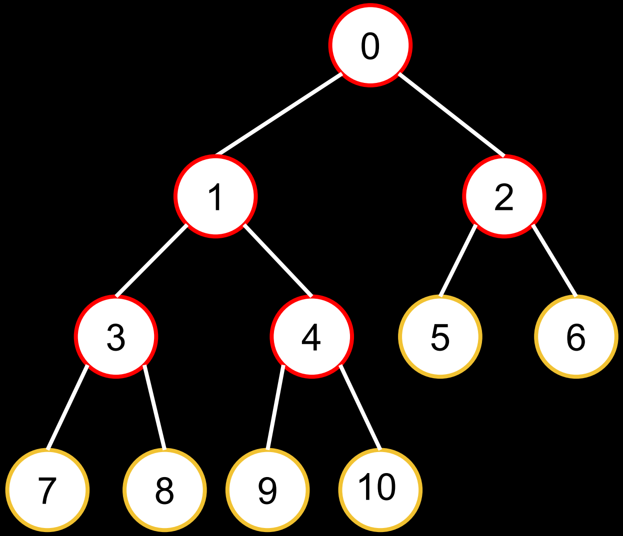
**Sethi** 1990 - Decision Trees as ANNs

1990 -> 2014 ???

**Welbl** 2014 - RFs as ANNs, refine trees

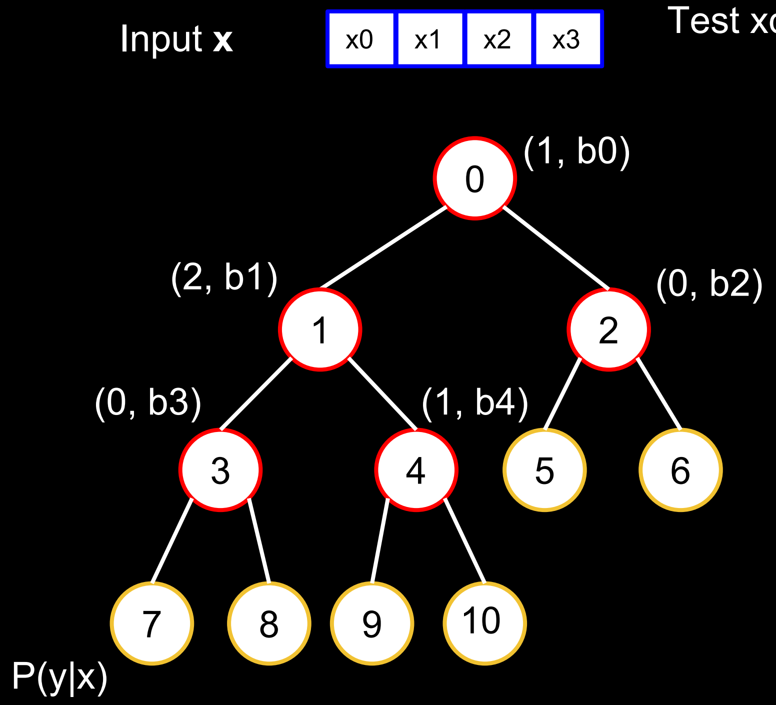
**Richmond** 2015 - Cascaded RFs as CNNs

# Decision Tree

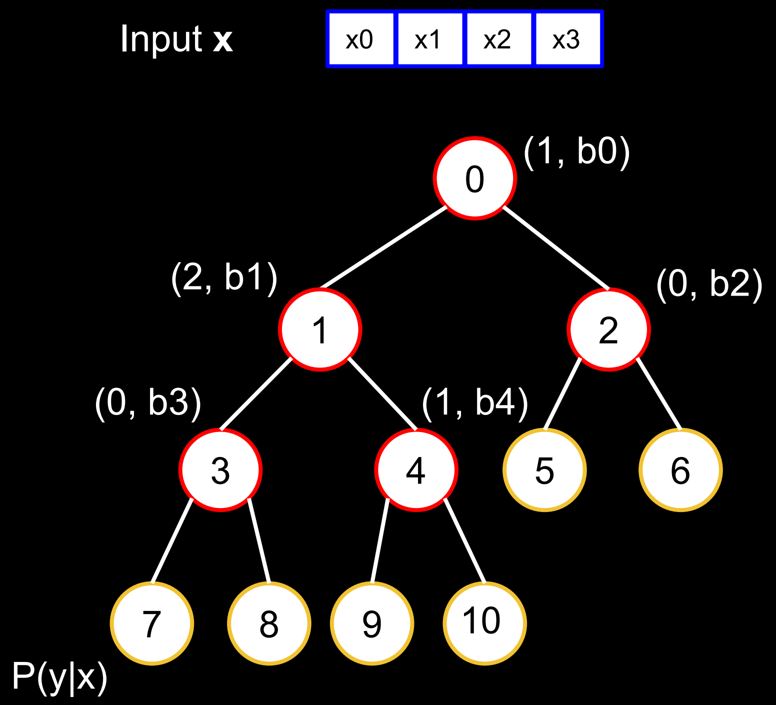


## Decision Tree

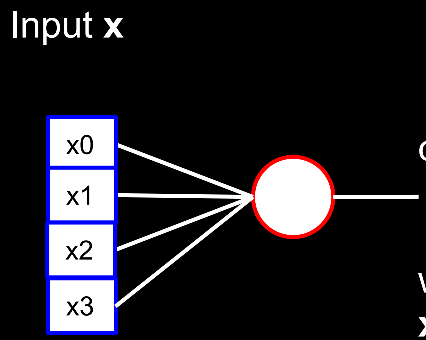
Here:
C num of classes = 2
D dimensionality = 4
Test xd >= bn



## Decision Tree



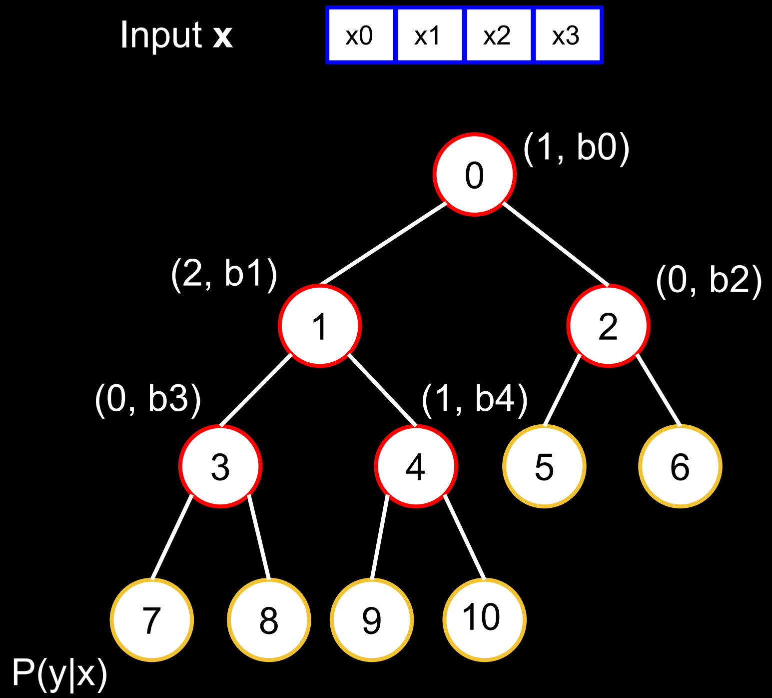

## Single Neuron



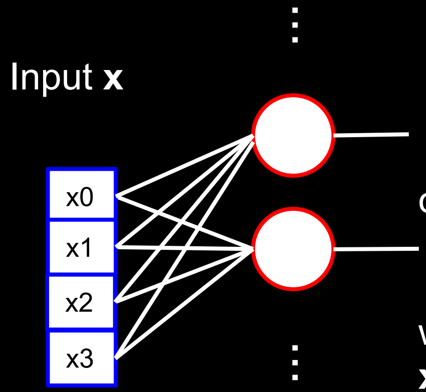

op = f(**xw** + b)

where
**x** is an 1x4 input
**w** is an 4x1 vector of weights
b is a 1x1 bias term
f is a nonlinear function

## Decision Tree




## Multiple Neurons

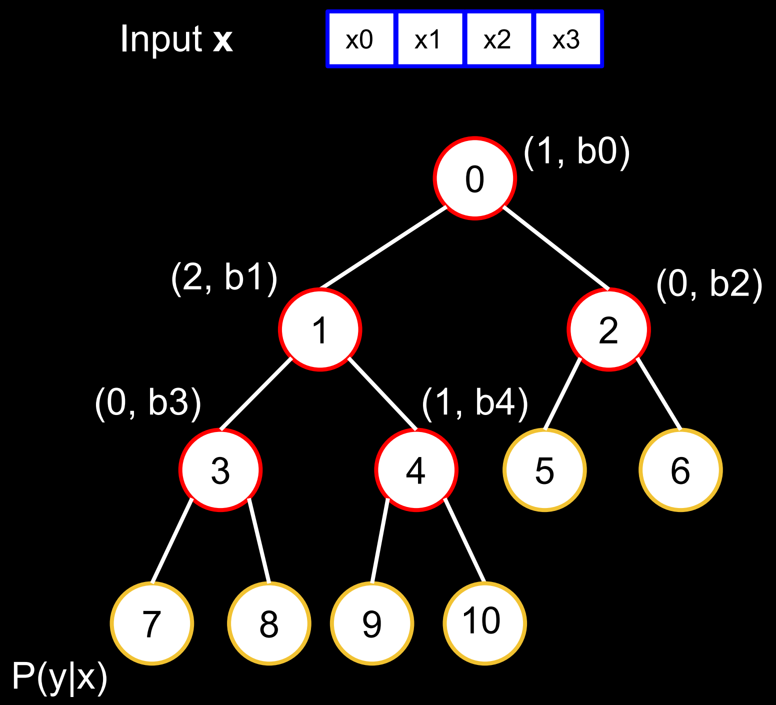



op = f(**xW** + **b**)

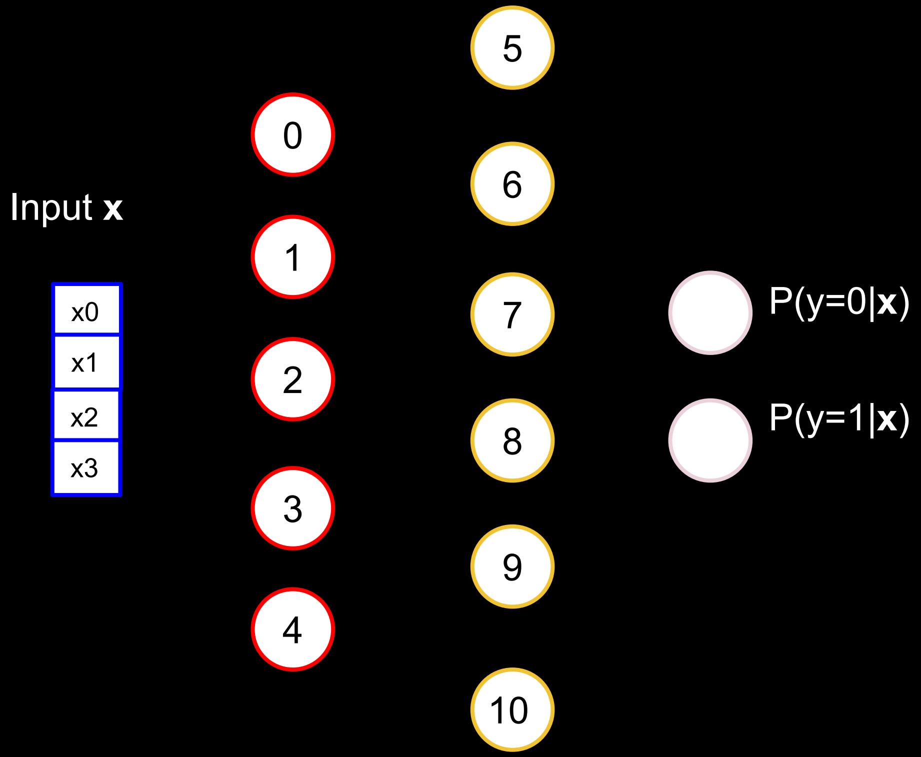
where
**x** is an 1x4 input
**W** is an 4x2 weight matrix
**b** is a 1x2 bias term
f is a nonlinear function

# Decision Tree as ANN

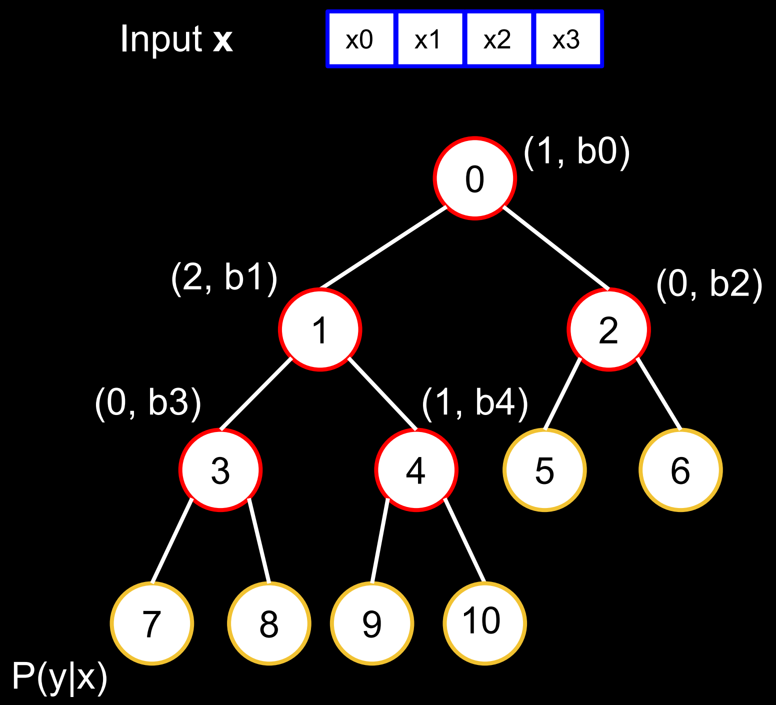
## Decision Tree



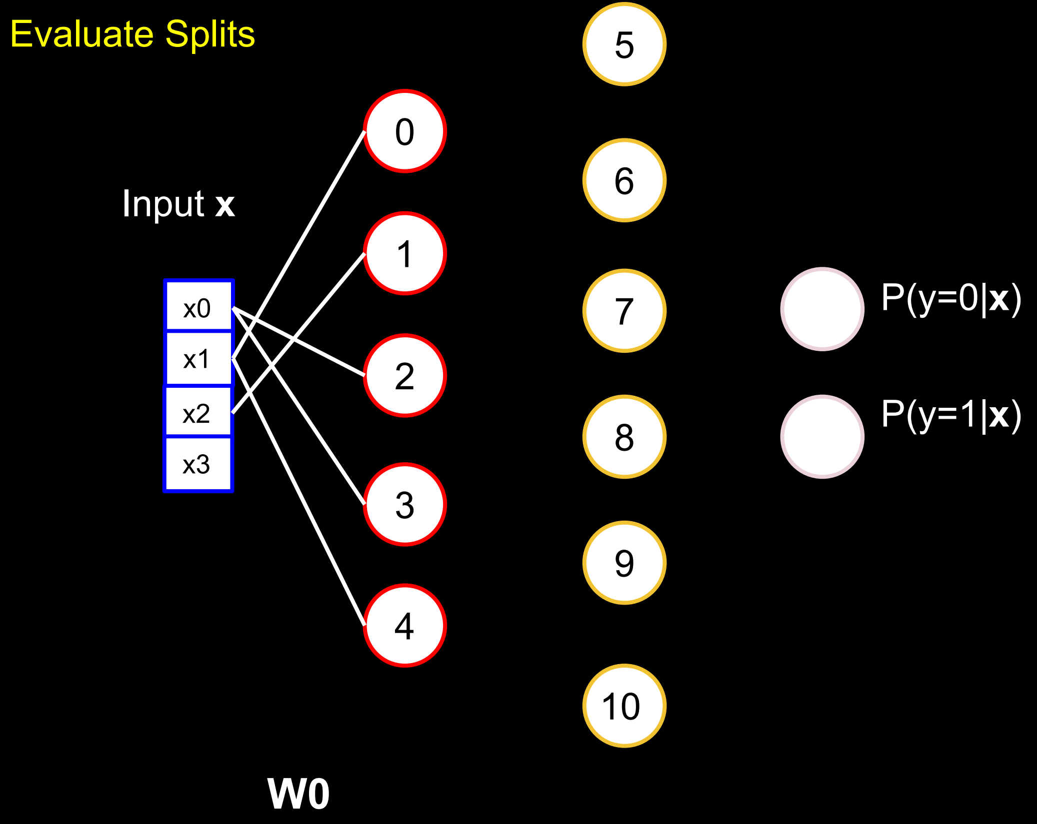

## Tree as ANN



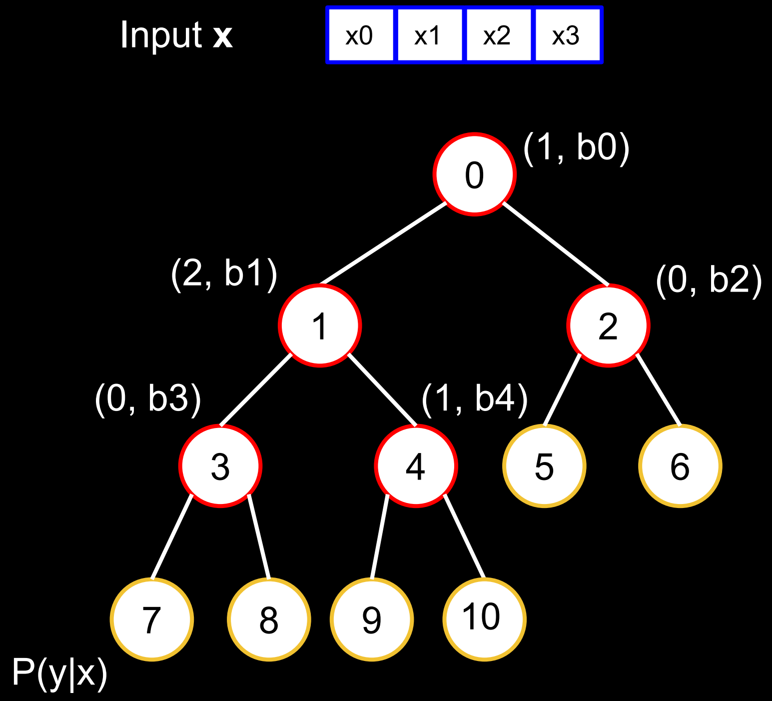
## Decision Tree



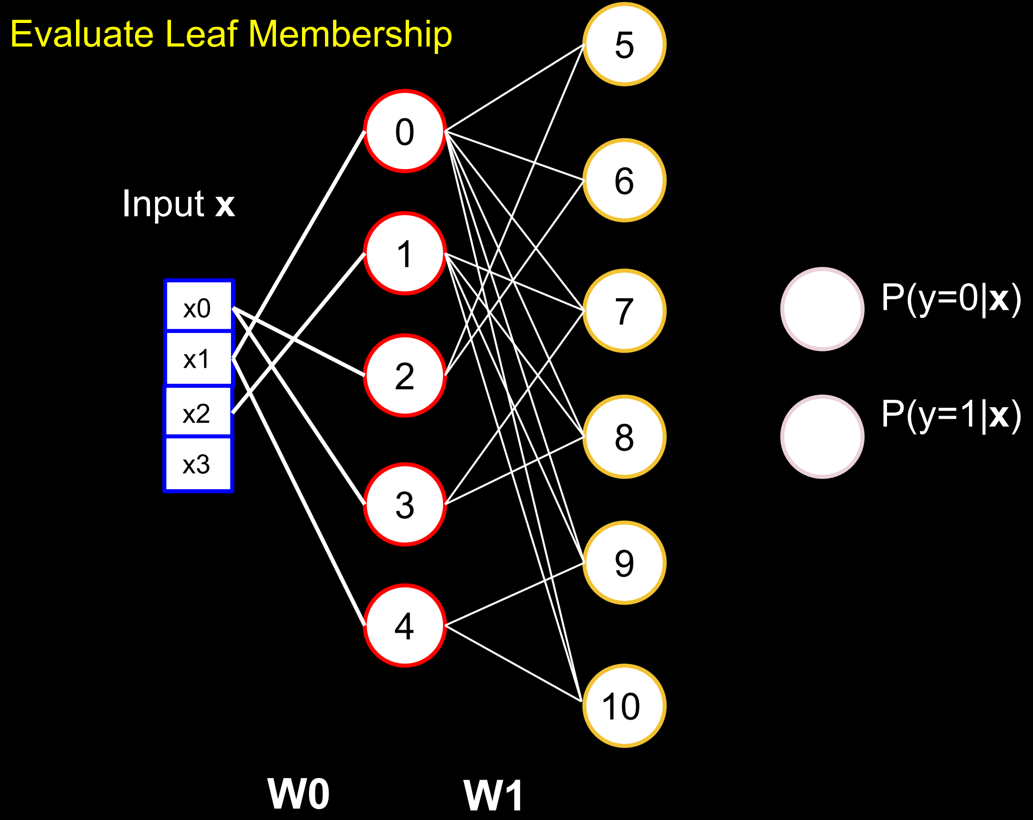

## Connect Input to Decision Nodes

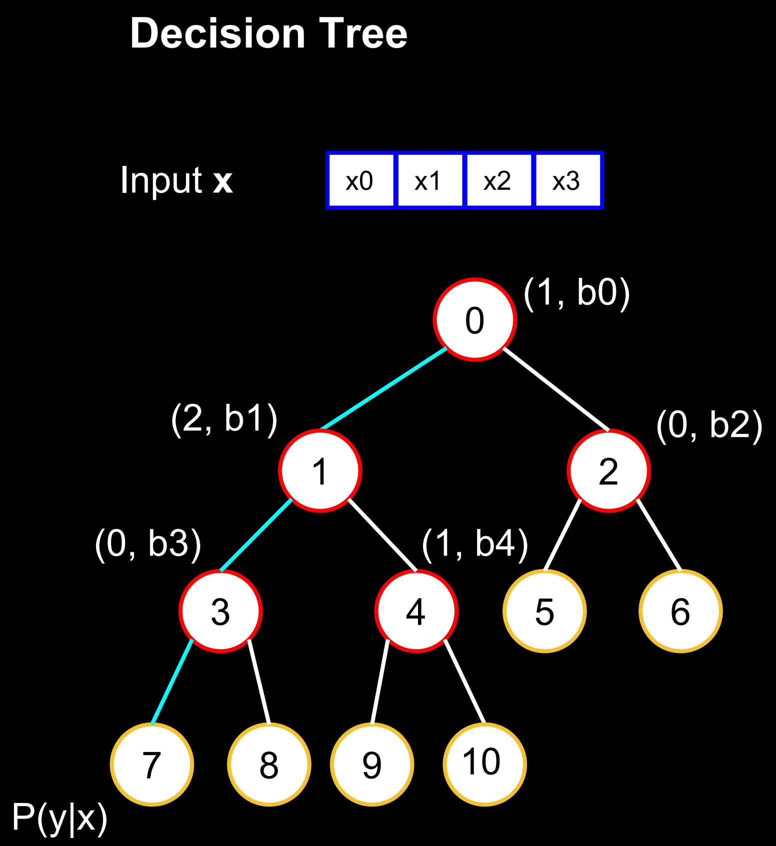


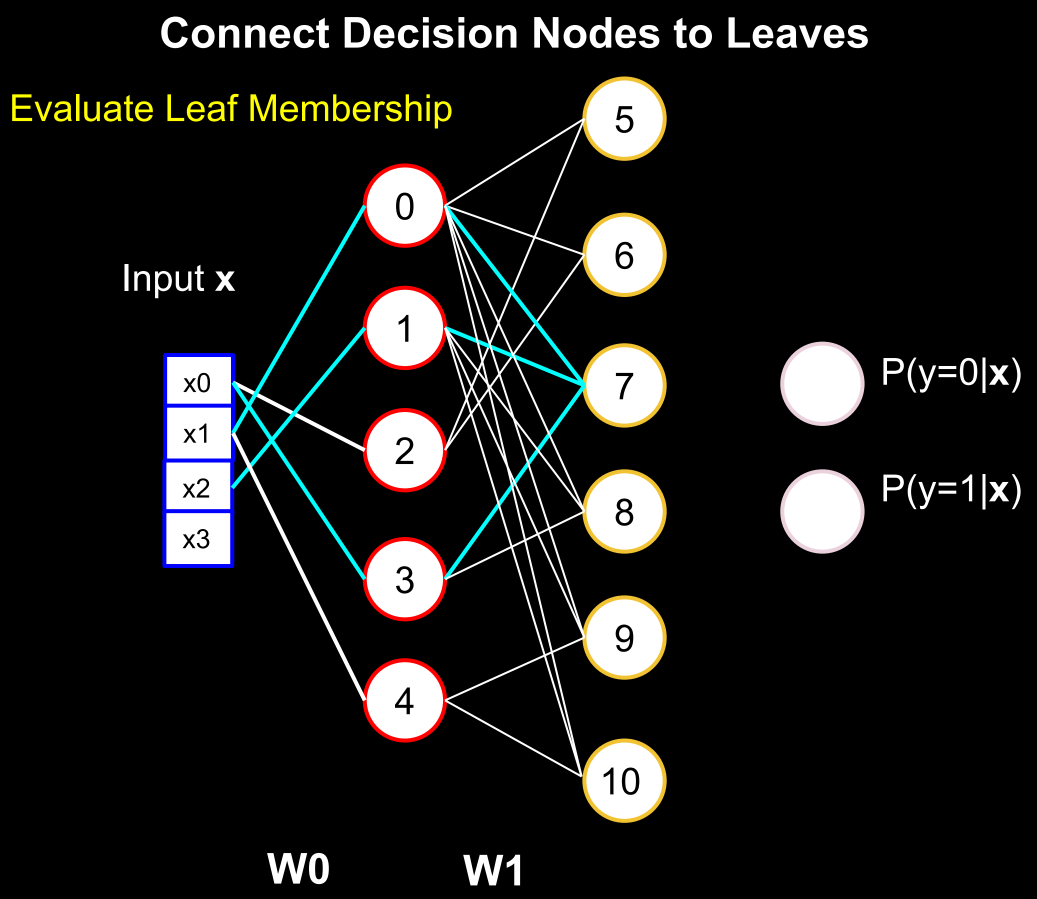
# Decision Tree



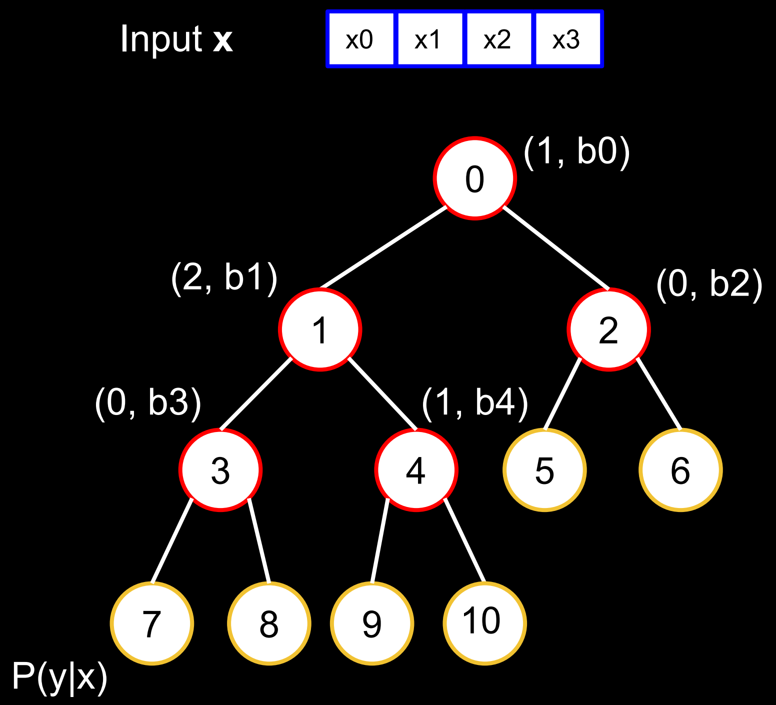

# Connect Decision Nodes to Leaves



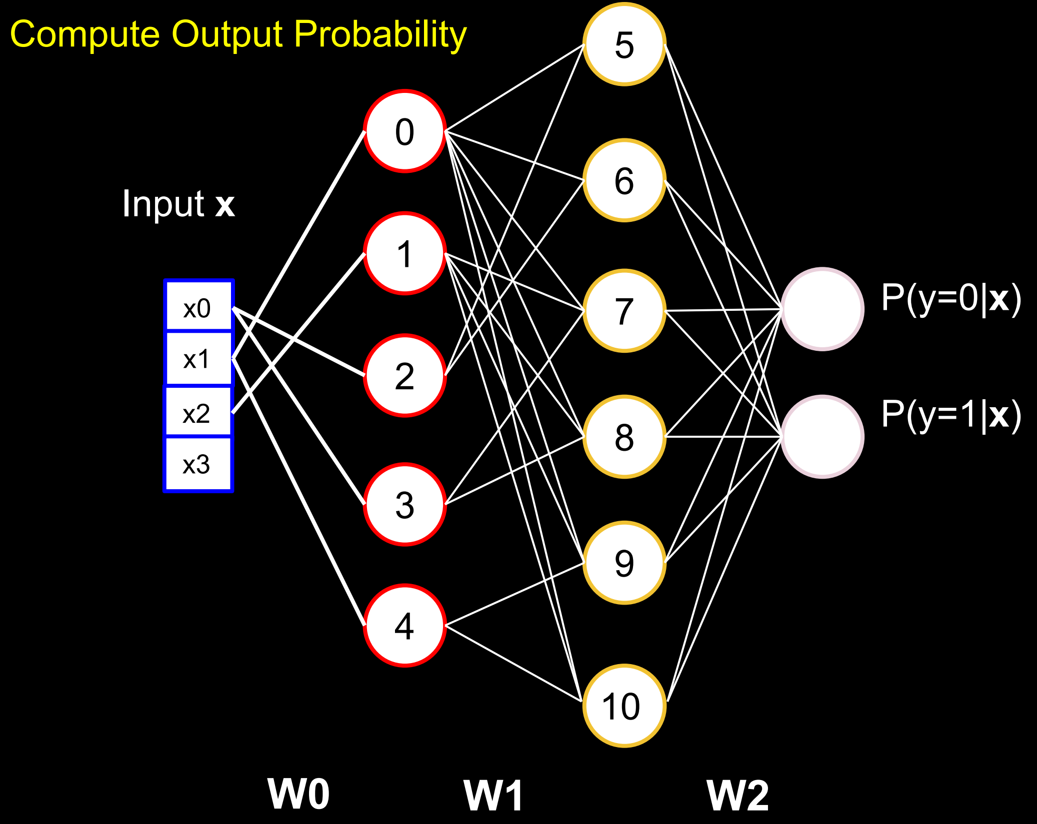

Decision Tree
Input **x**
x0
x1
x2
x3
0
(1, b0)
(2, b1)
1
2
(0, b2)
(0, b3)
(1, b4)
3
4
5
6
7
8
9
10
P(y|x)


Connect Decision Nodes to Leaves
Evaluate Leaf Membership
Input **x**
x0
x1
x2
x3
0
1
2
3
4
5
6
7
8
9
10
W0
W1
P(y=0|**x**)
P(y=1|**x**)

## Decision Tree



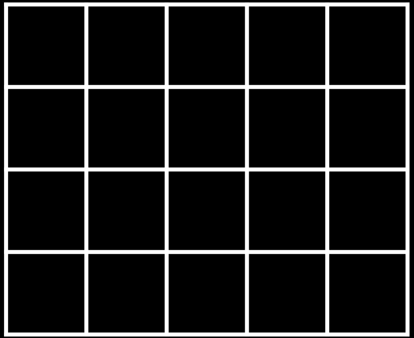

## Connect Leaves to Output



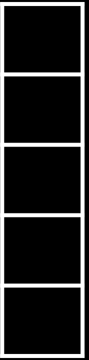
# Network Parameters

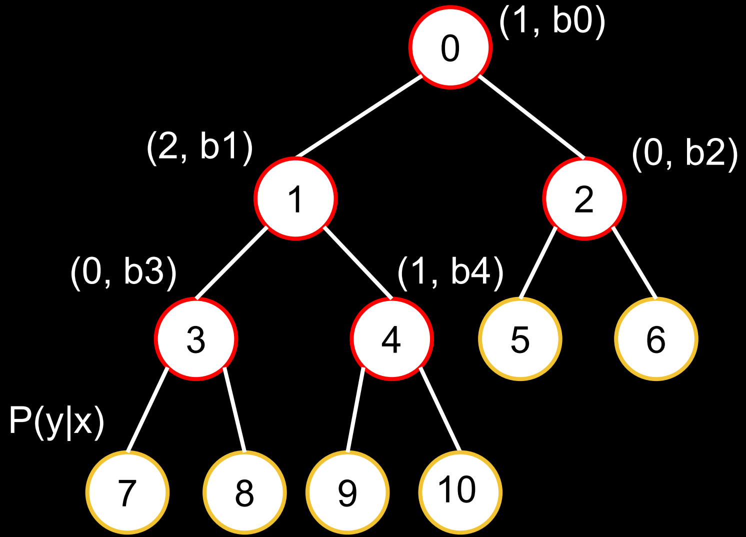
## Connect Input to Decision Nodes

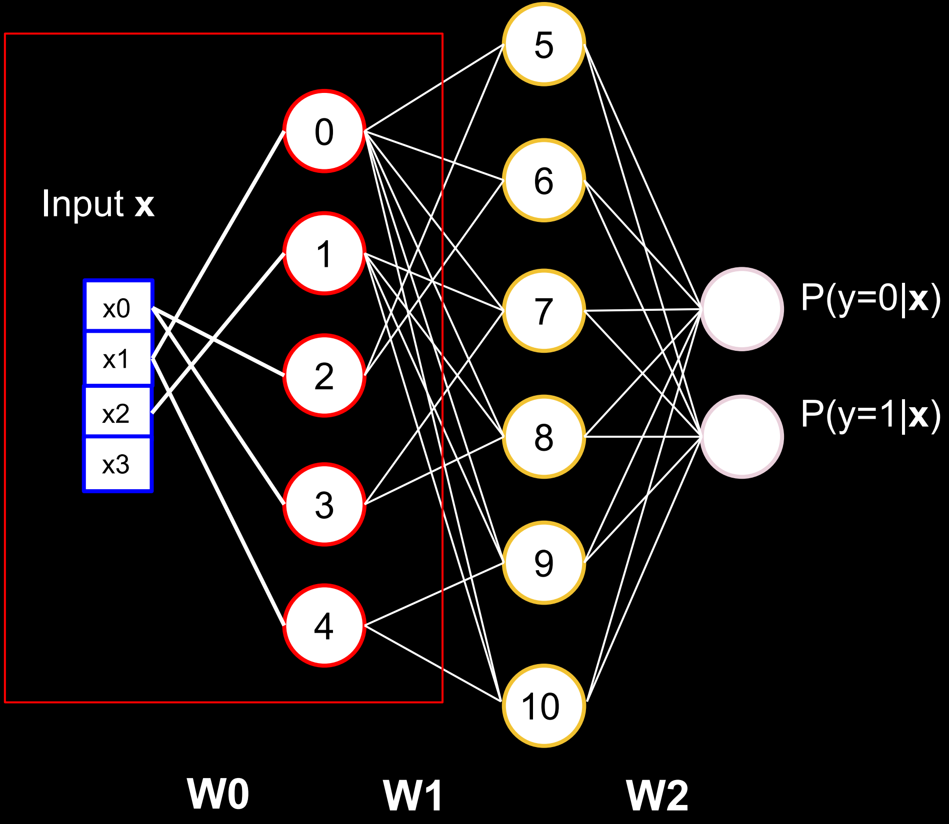
**W0 =**



**b0 =**



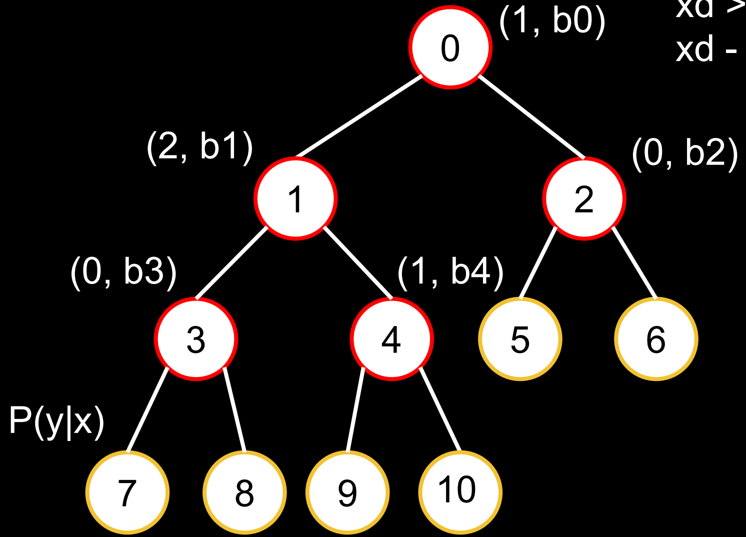


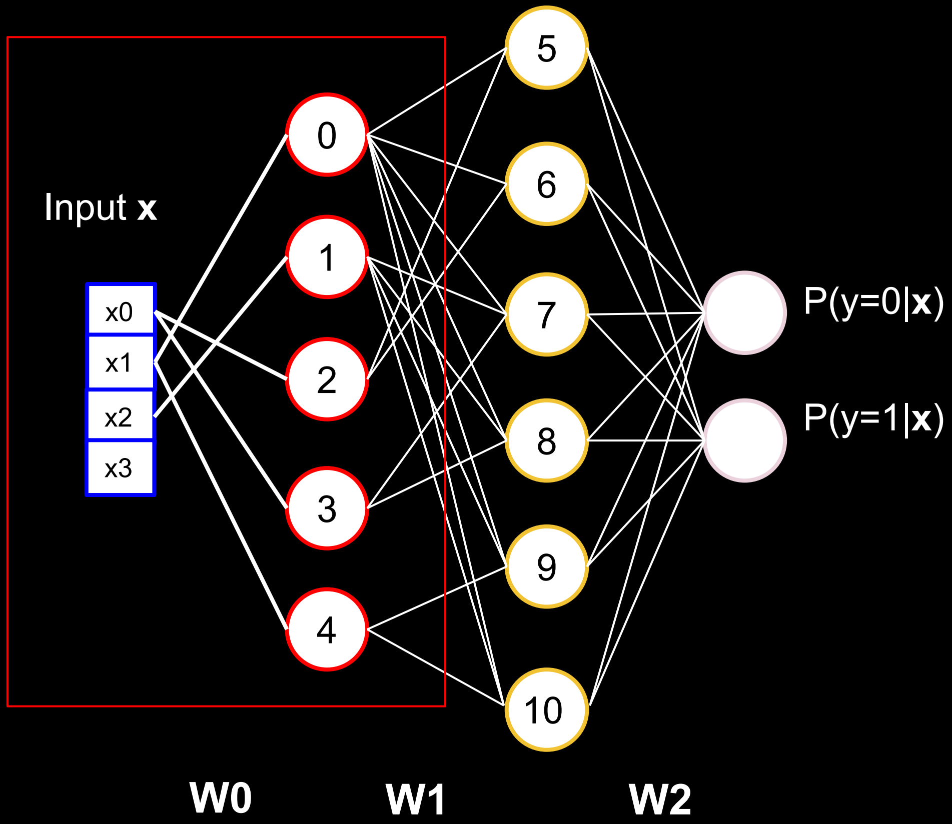



## Connect Input to Decision Nodes

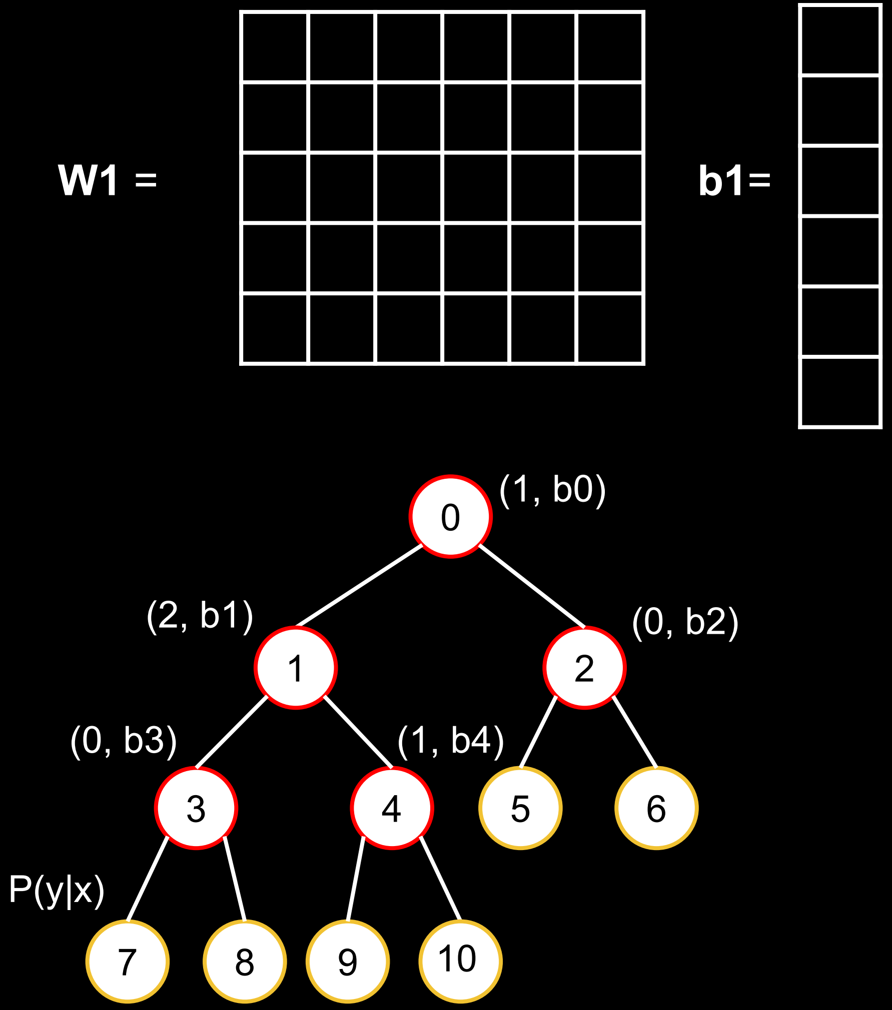
$$\mathbf{W0} = \begin{bmatrix} & & 1 & 1 & \\ 1 & & & & 1 \\ & 1 & & & \\ & & & & \end{bmatrix} \qquad \mathbf{b0} = \begin{bmatrix} -b0 \\ -b1 \\ -b2 \\ -b3 \\ -b4 \end{bmatrix}$$

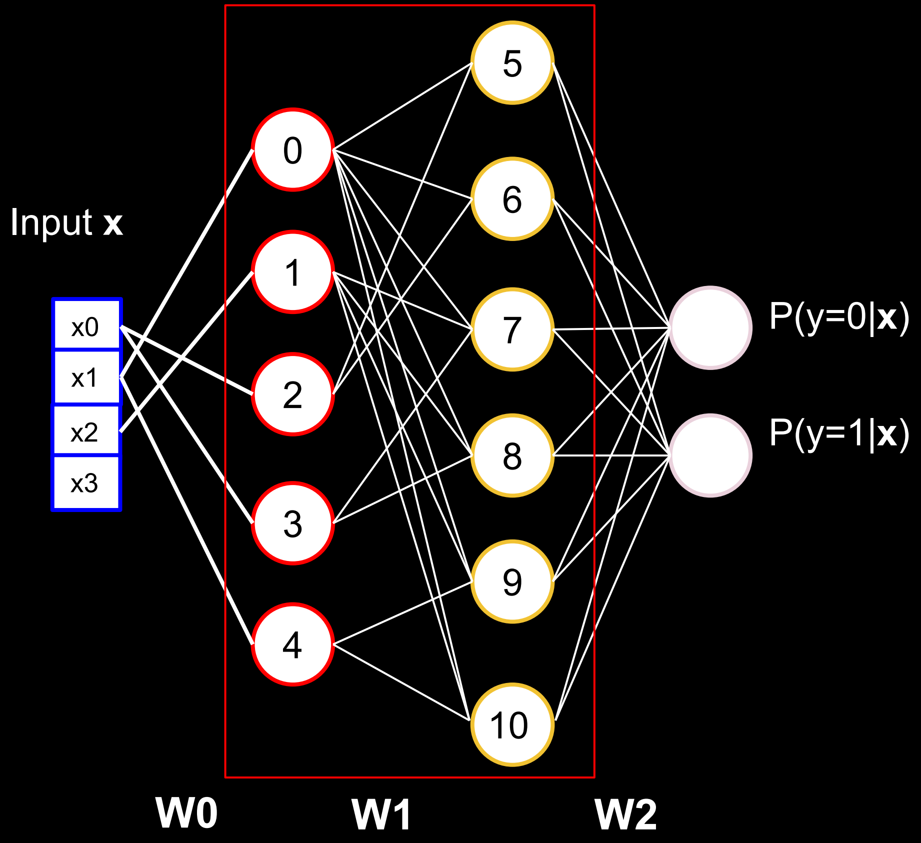
Node Test:
xd >= bn
xd - bn >= 0



# Connect Decision Nodes to Leaves



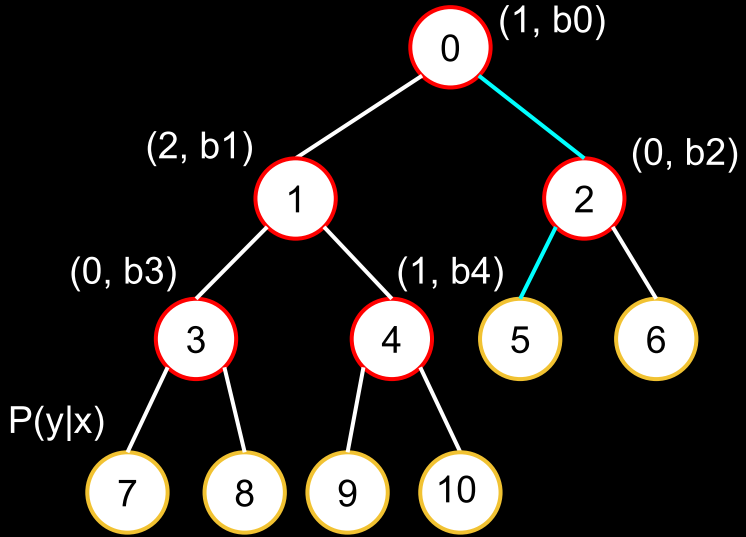
## Connect Decision Nodes to Leaves

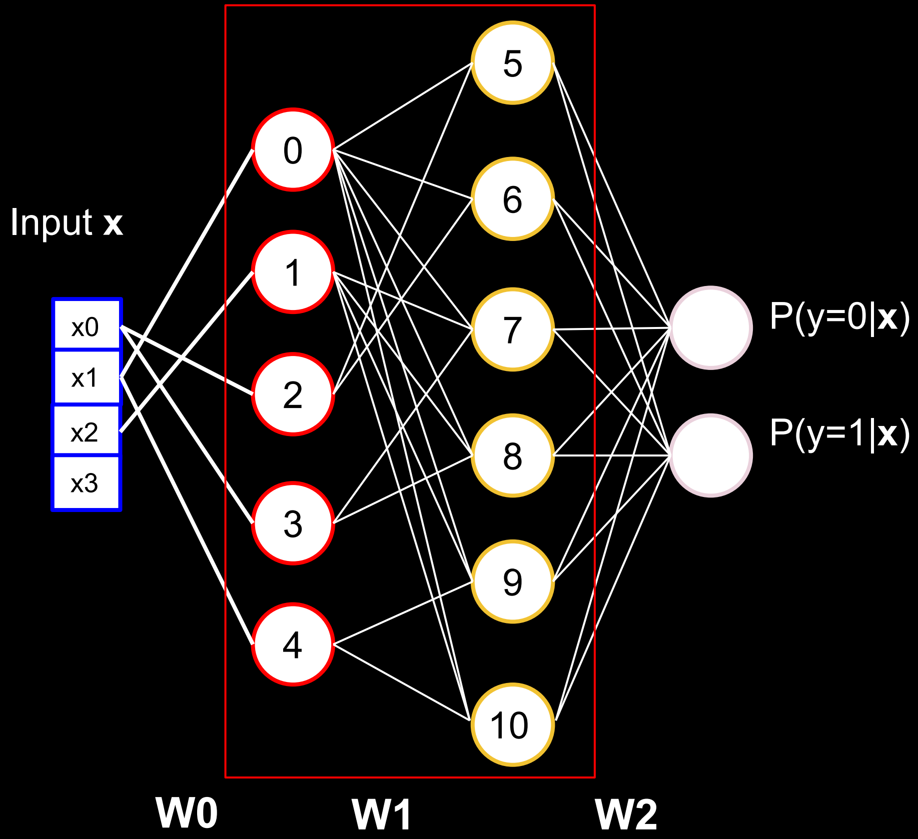
**W1** =

| | | | | | |
|---|---|---|---|---|---|
| 1 | | | | | |
| | | | | | |
| -1 | | | | | |
| | | | | | |
| | | | | | |

**b1**=

| |
|---|
| |
| |
| |
| |
| |
| |



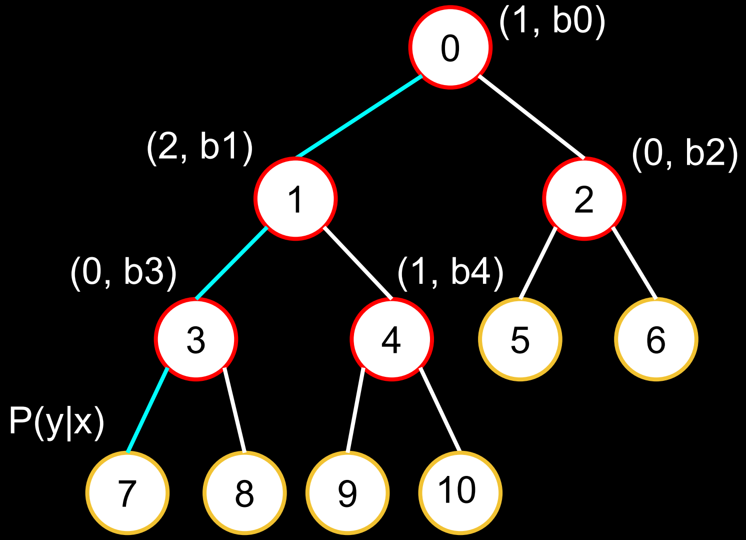
**W1 =**

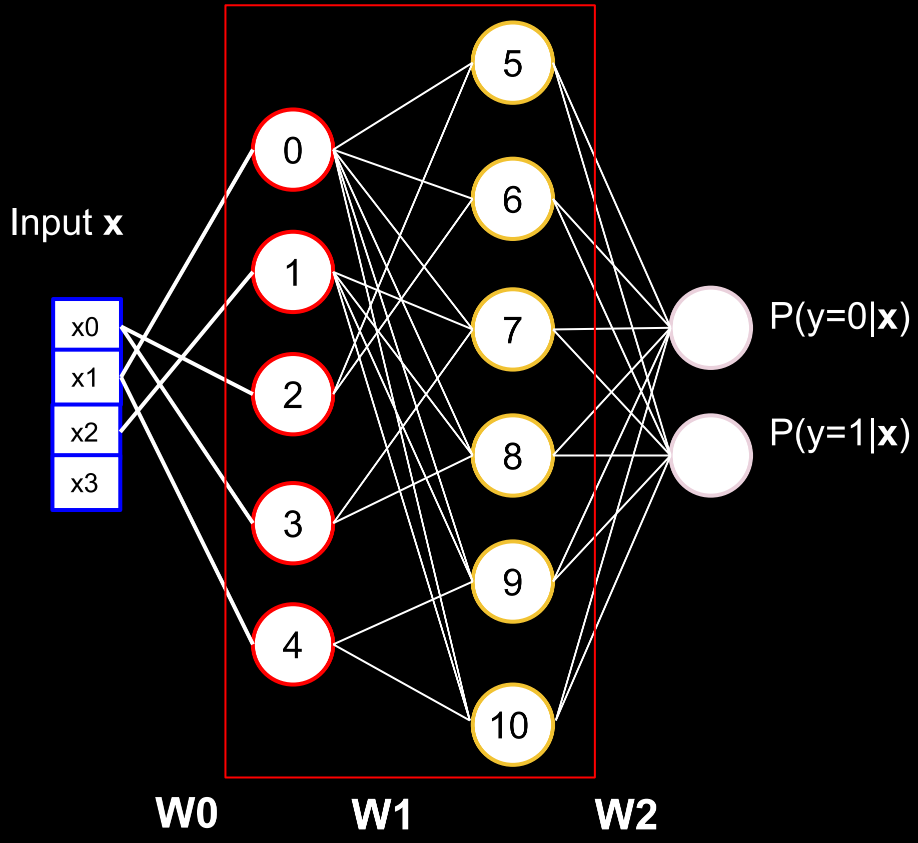
| | | | | | |
|---|---|---|---|---|---|
| 1 | | -1 | | | |
| | | -1 | | | |
| -1 | | | | | |
| | | -1 | | | |
| | | | | | |

**b1=**

| |
|---|
| |
| |
| |
| |
| |
| |

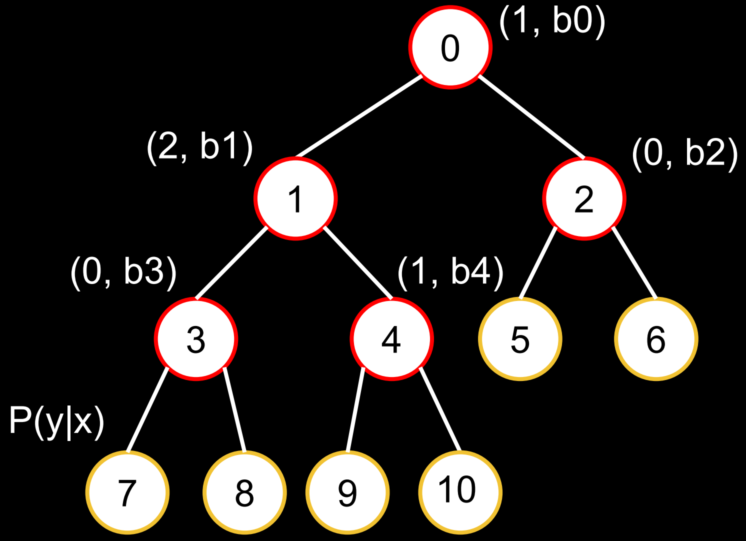
## Connect Decision Nodes to Leaves

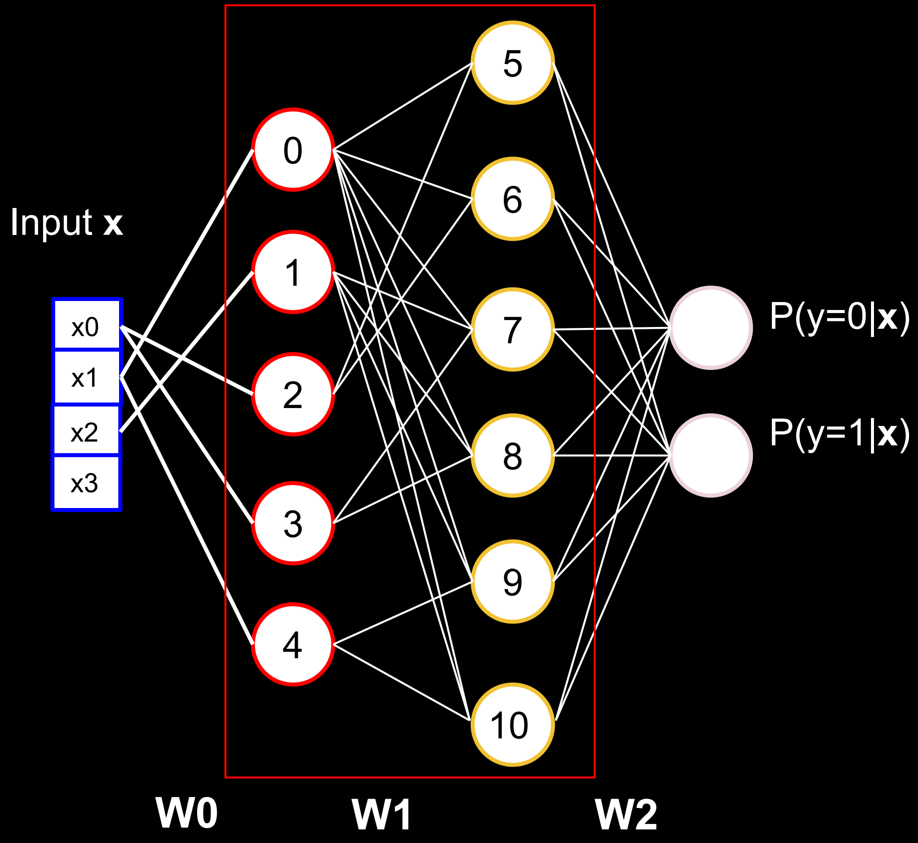


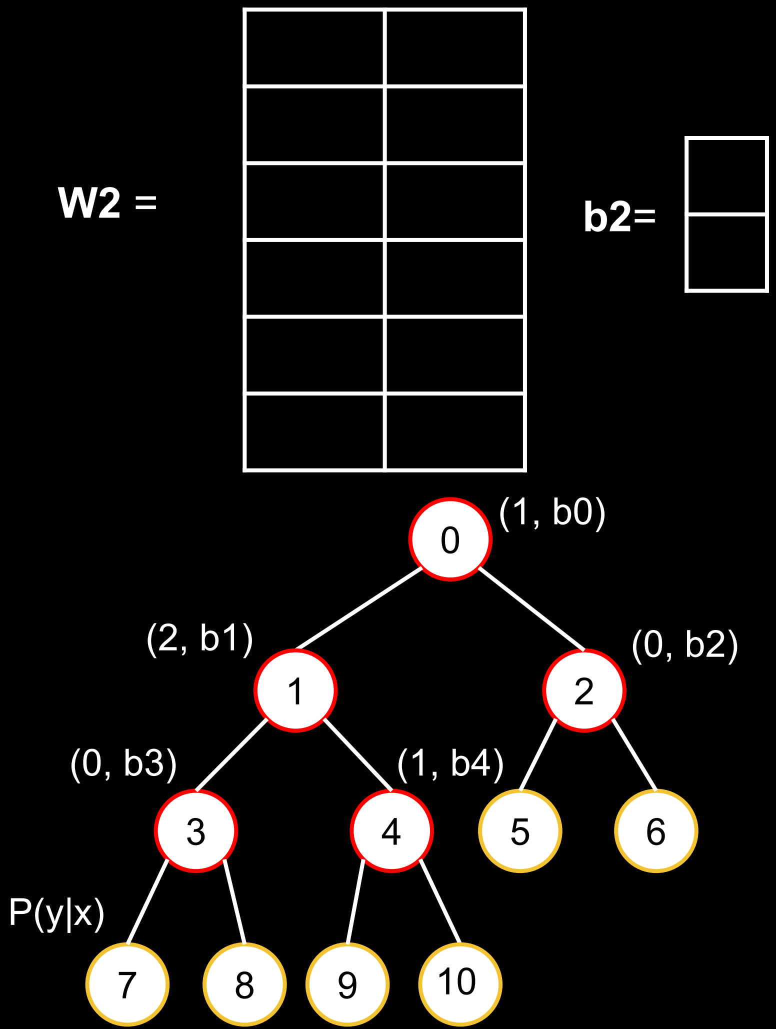
## Connect Decision Nodes to Leaves

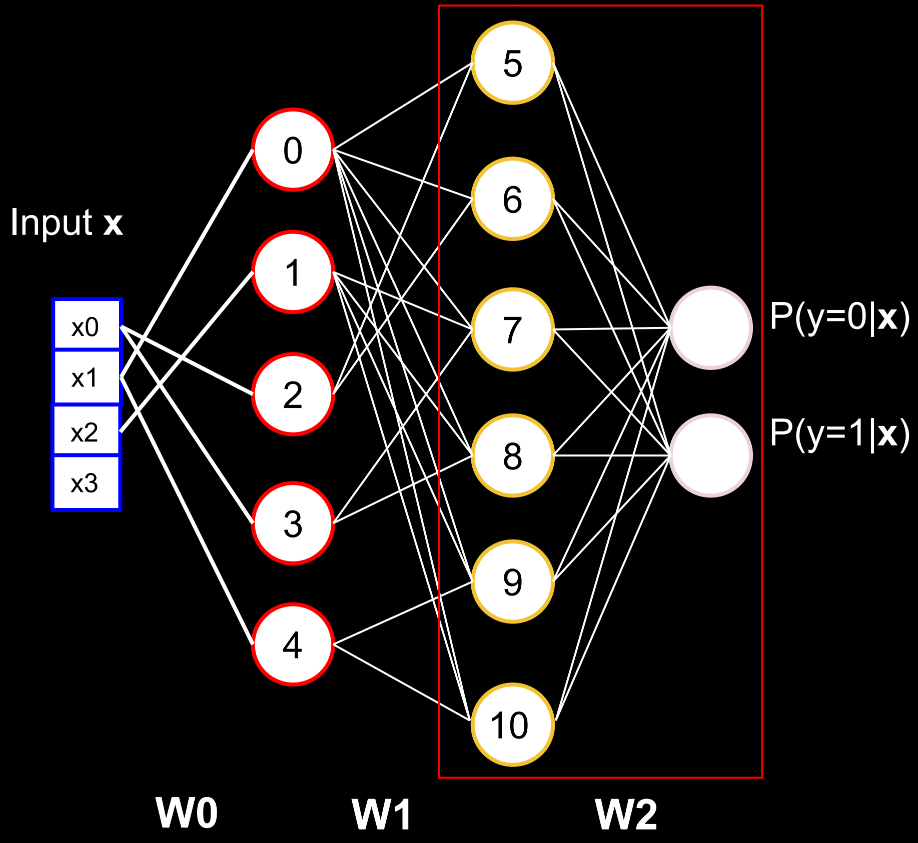
$$\mathbf{W1} = \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 \\ & & -1 & -1 & 1 & 1 \\ -1 & 1 & & & & \\ & & -1 & 1 & & \\ & & & & -1 & 1 \end{bmatrix} \qquad \mathbf{b1} = \begin{bmatrix} -2 \\ -2 \\ -3 \\ -3 \\ -3 \\ -3 \end{bmatrix}$$



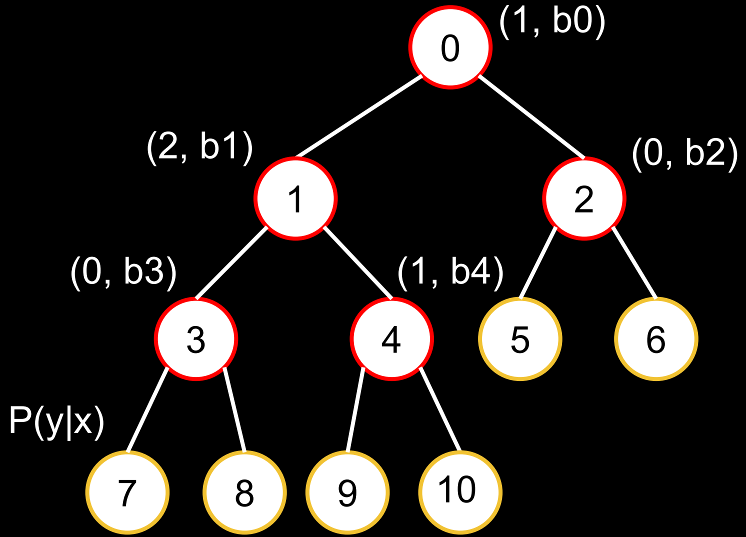
# Connect Leaves to Output

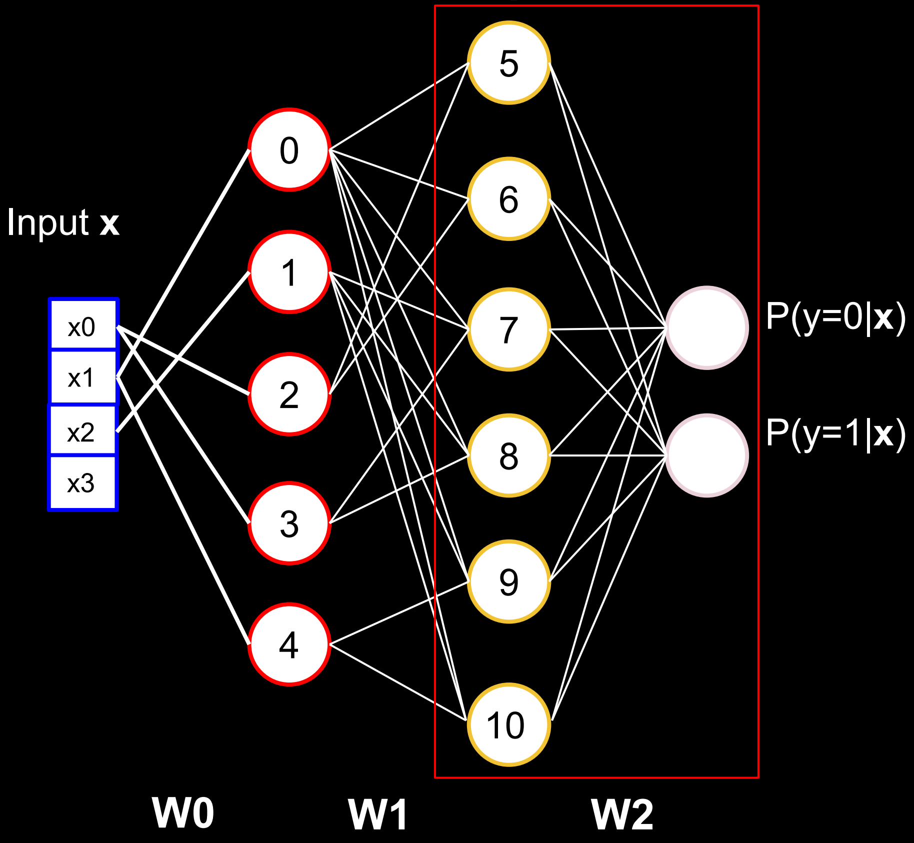


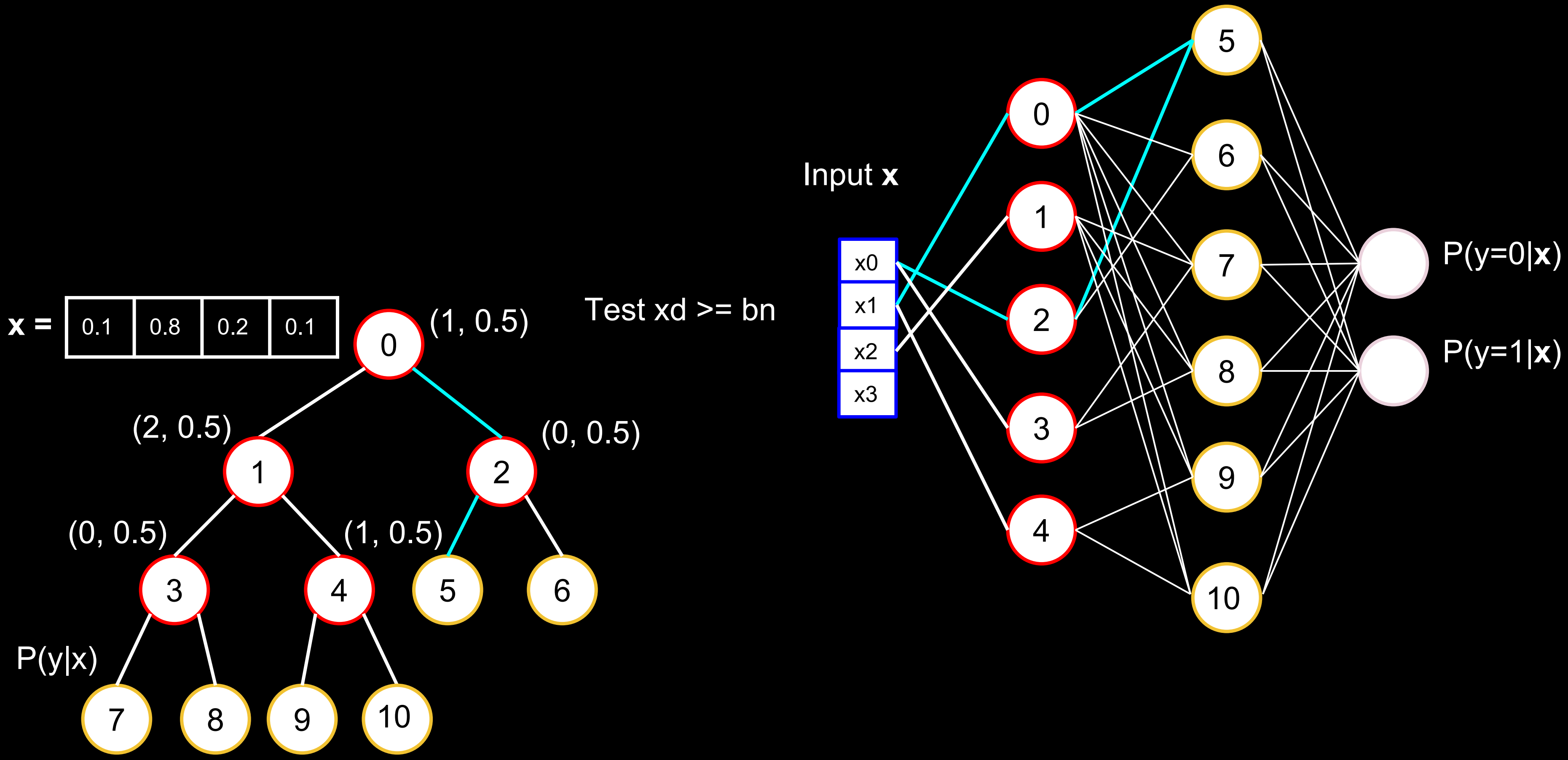
# Connect Leaves to Output

$$
\mathbf{W2} = \begin{bmatrix} P_5(y{==}0) & P_5(y{==}1) \\ P_6(y{==}0) & P_6(y{==}1) \\ P_7(y{==}0) & P_7(y{==}1) \\ P_8(y{==}0) & P_8(y{==}1) \\ P_9(y{==}0) & P_9(y{==}1) \\ P_{10}(y{==}0) & P_{10}(y{==}1) \end{bmatrix}
\qquad
\mathbf{b2} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
$$



# Example Feed Forward Pass

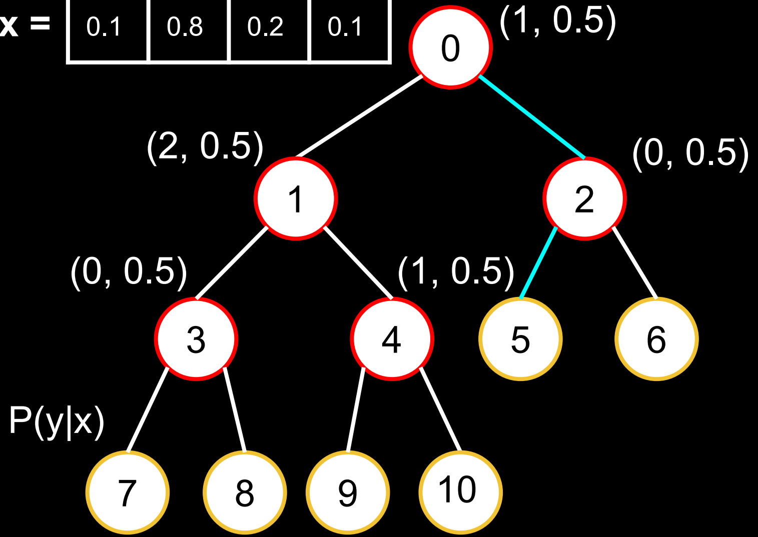


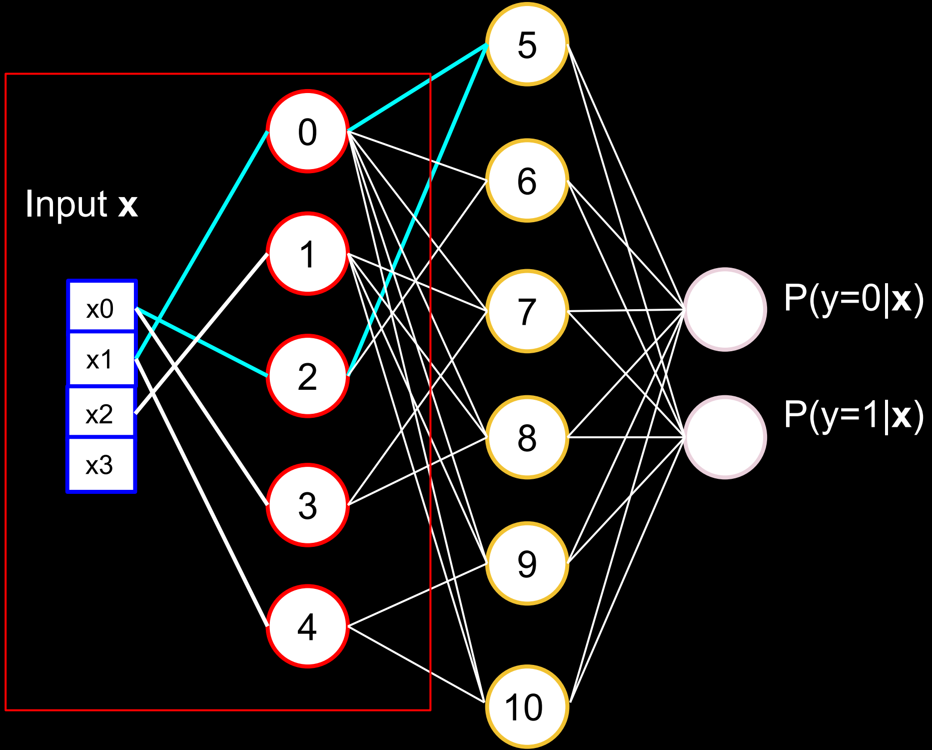
x =
0.1
0.8
0.2
0.1
(1, 0.5)
(2, 0.5)
(0, 0.5)
(0, 0.5)
(1, 0.5)
P(y|x)
0
1
2
3
4
5
6
7
8
9
10
Test xd >= bn
Input **x**
x0
x1
x2
x3
P(y=0|**x**)
P(y=1|**x**)

$$\begin{bmatrix}0.1\\0.8\\0.2\\0.1\end{bmatrix} \quad \begin{bmatrix}0 & 0 & 1 & 1 & 0\\1 & 0 & 0 & 0 & 1\\0 & 1 & 0 & 0 & 0\\0 & 0 & 0 & 0 & 0\end{bmatrix} + \begin{bmatrix}-0.5\\-0.5\\-0.5\\-0.5\\-0.5\end{bmatrix}$$

op0 = f(**xW0** + **b0**)

**x** = $\begin{bmatrix}0.1 & 0.8 & 0.2 & 0.1\end{bmatrix}$

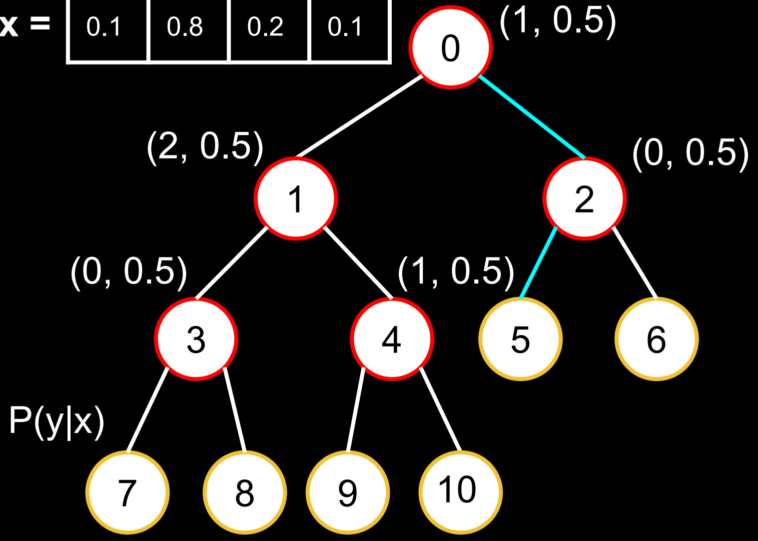


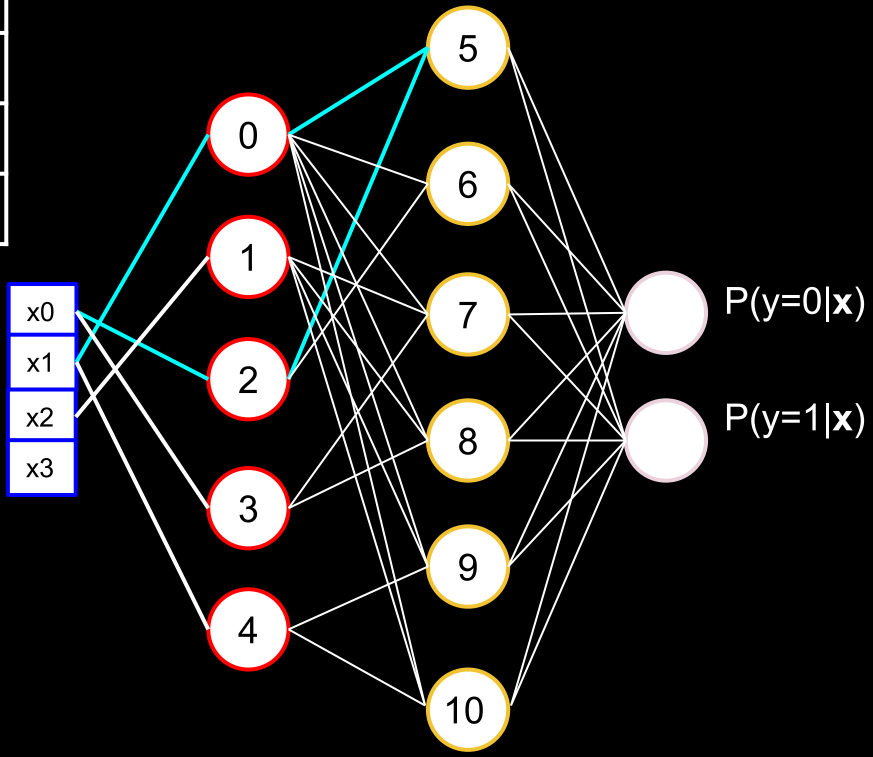



$$\begin{bmatrix} 0.1 \\ 0.8 \\ 0.2 \\ 0.1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} -0.5 \\ -0.5 \\ -0.5 \\ -0.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 0.3 \\ -0.3 \\ -0.4 \\ -0.4 \\ 0.3 \end{bmatrix}$$

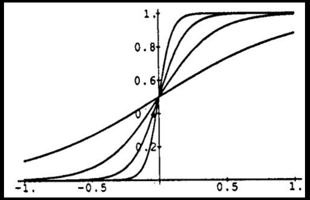
**xW0 + b0**

**x** = | 0.1 | 0.8 | 0.2 | 0.1 |



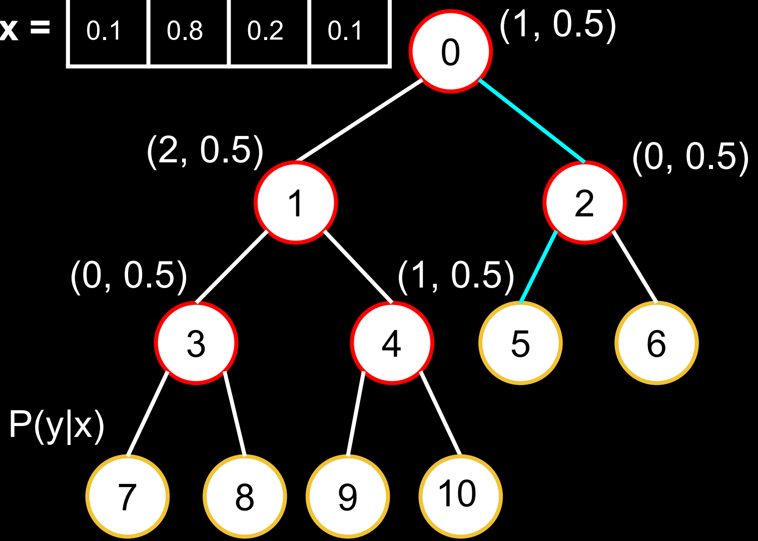
$$f\left(\begin{bmatrix}0.3\\-0.3\\-0.4\\-0.4\\0.3\end{bmatrix}\right)=\begin{bmatrix}1\\-1\\-1\\-1\\1\end{bmatrix}$$

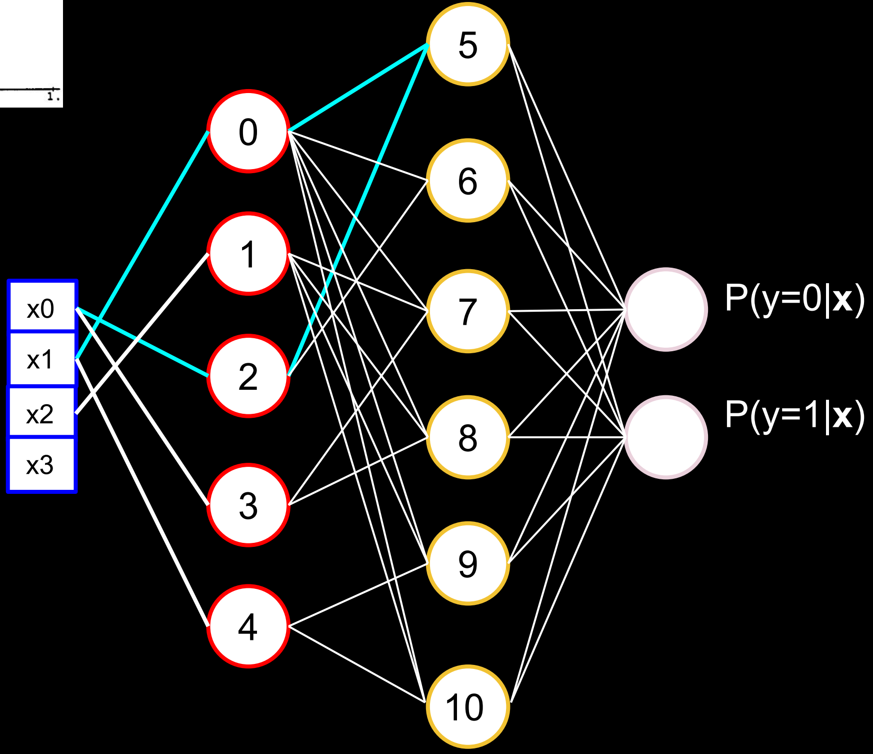



< 0 : -1
>=0 : 1

**op0** = f(**xW0** + **b0**)

**x** = | 0.1 | 0.8 | 0.2 | 0.1 |



| |
|---|
| 1 |
| -1 |
| -1 |
| -1 |
| 1 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | 1 | -1 | -1 | -1 | -1 |
| | | -1 | -1 | 1 | 1 |
| -1 | 1 | | | | |
| | | -1 | 1 | | |
| | | | | -1 | 1 |

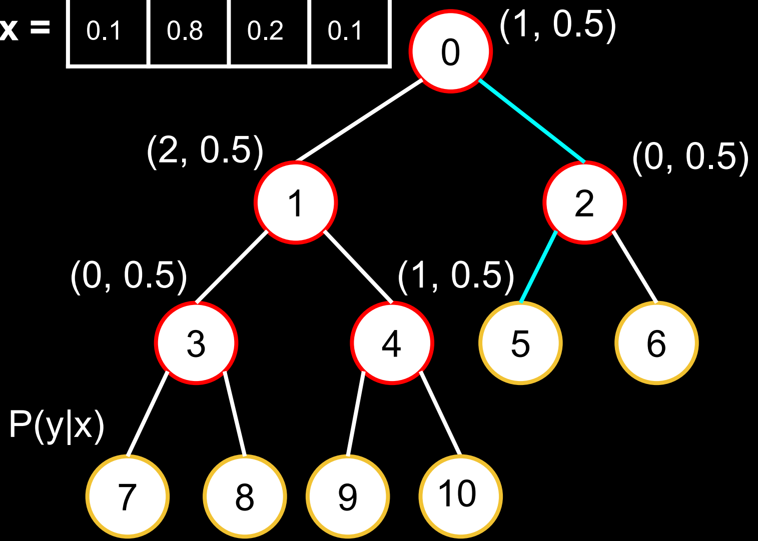
+

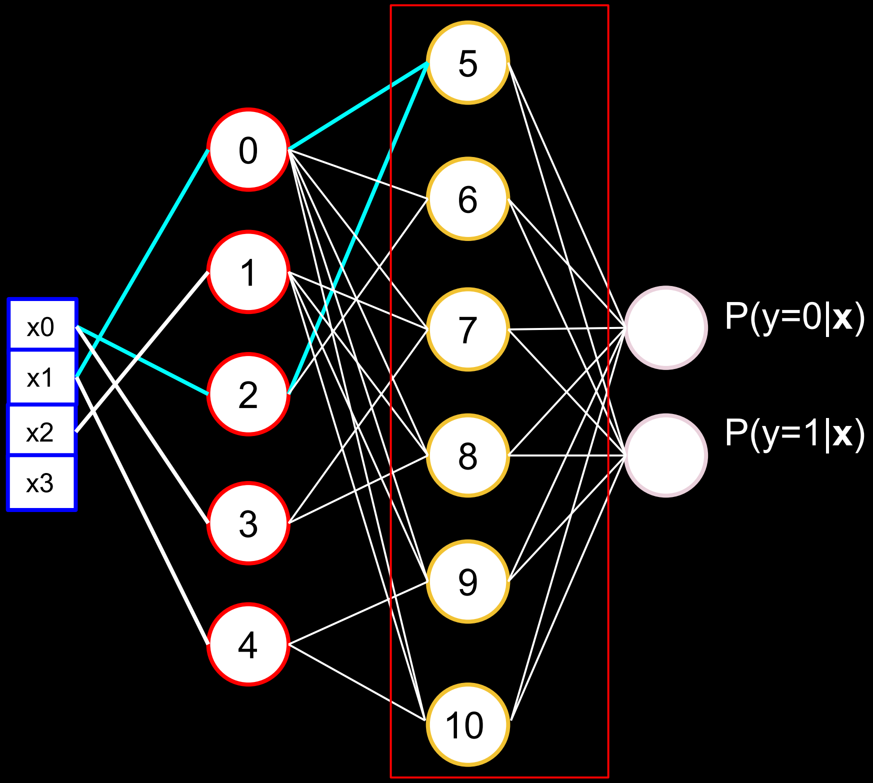
| |
|---|
| -2 |
| -2 |
| -3 |
| -3 |
| -3 |
| -3 |

**op1** = g(**op0W1** + **b1**)

**x** =

| | | | |
|---|---|---|---|
| 0.1 | 0.8 | 0.2 | 0.1 |



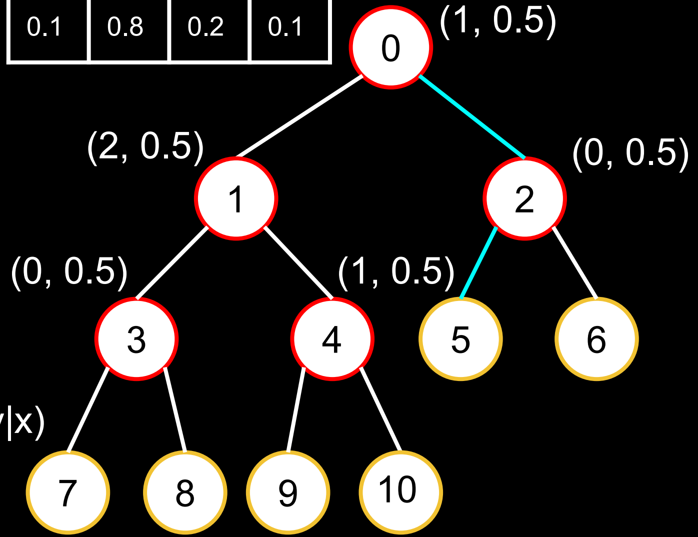



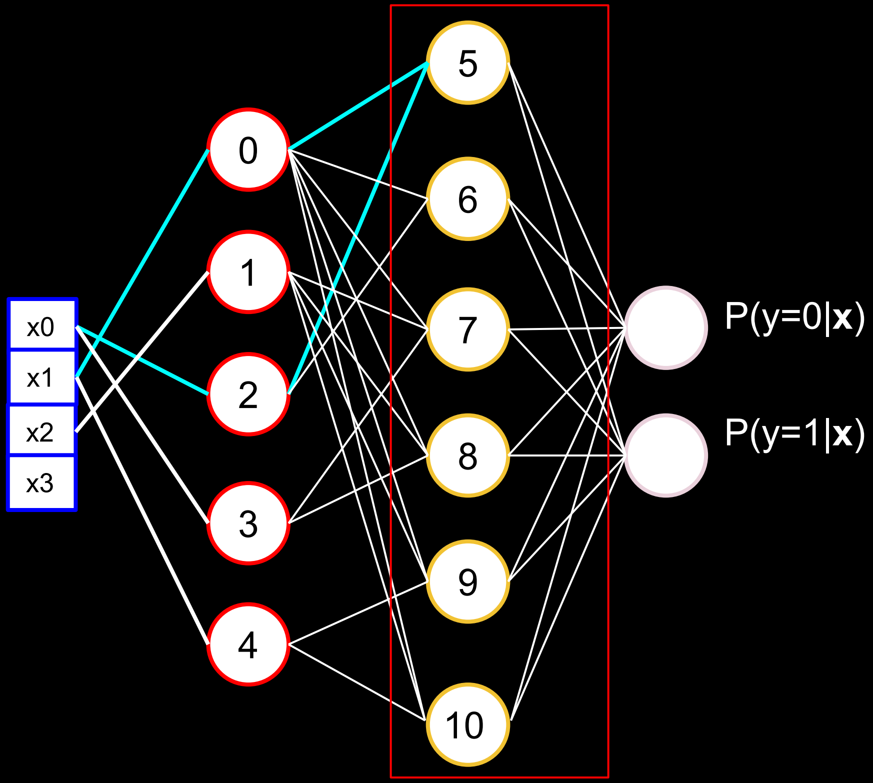
$$\begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 \\ & & -1 & -1 & 1 & 1 \\ -1 & 1 & & & & \\ & & -1 & 1 & & \\ & & & & -1 & 1 \end{bmatrix} + \begin{bmatrix} -2 \\ -2 \\ -3 \\ -3 \\ -3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ -2 \\ -4 \\ -6 \\ -4 \end{bmatrix}$$

**op0W1 + b1**

**x** = 

| | | | |
|---|---|---|---|
| 0.1 | 0.8 | 0.2 | 0.1 |



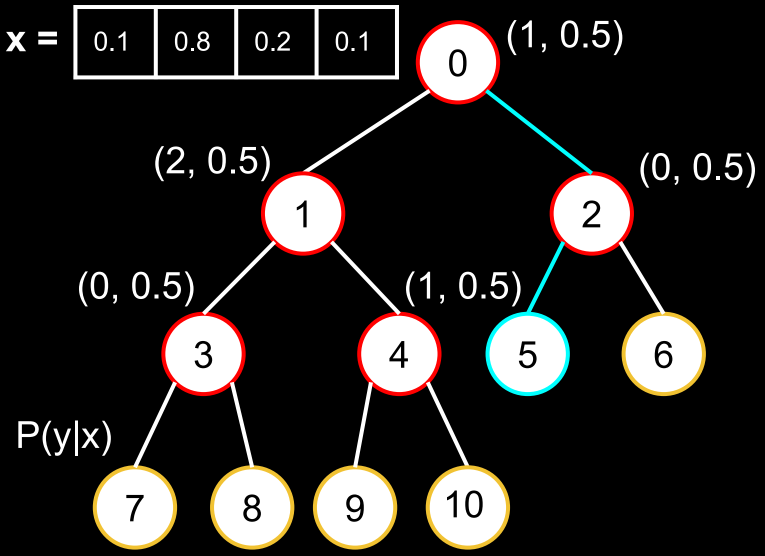



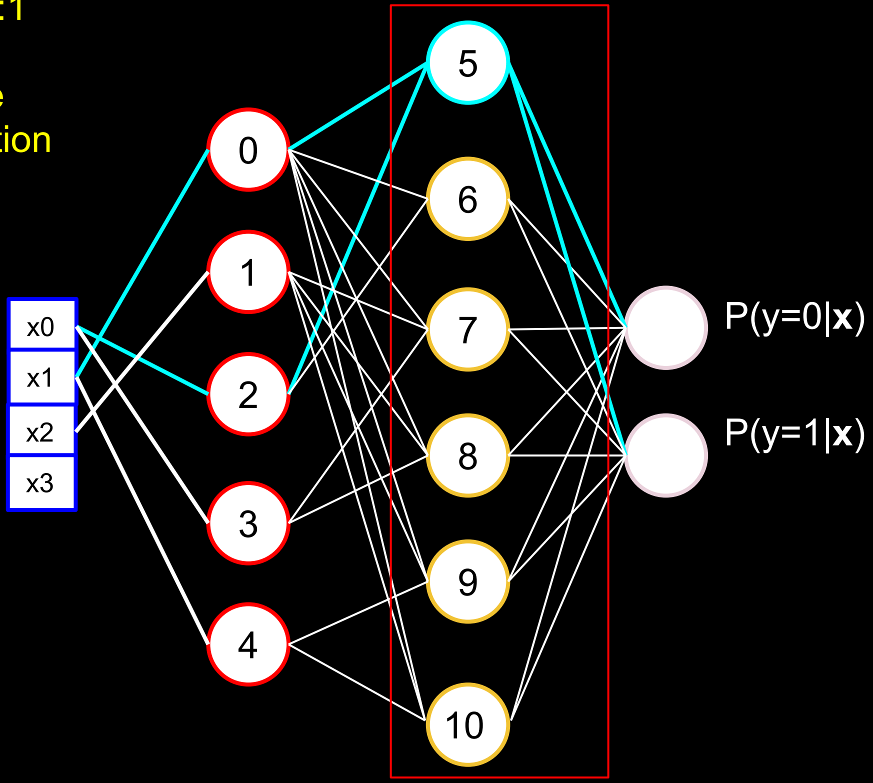
$$g\left(\begin{bmatrix} 0 \\ -2 \\ -2 \\ -4 \\ -6 \\ -4 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

**op1 = g(op0W1 + b1)**

Only want this in range 0:1, not -1:1

Result will be the maximum activation (==0)



| | | | | | |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 |

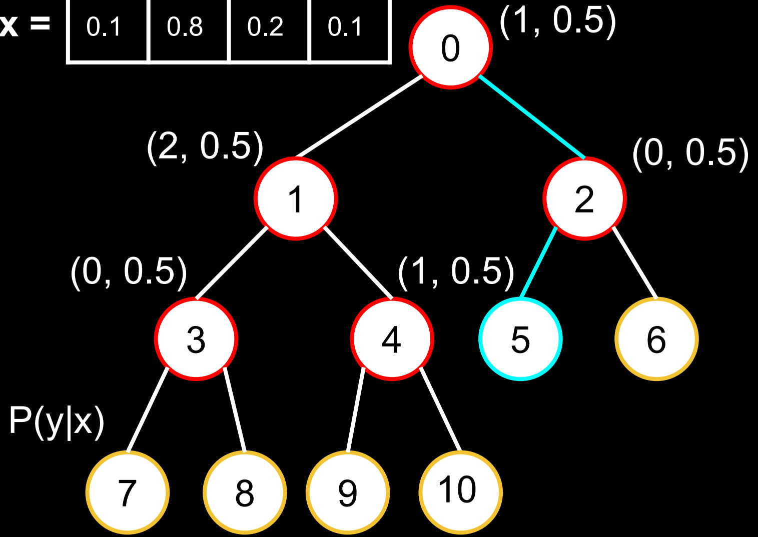
| | |
|---|---|
| $P_5(y==0)$ | $P_5(y==1)$ |
| $P_6(y==0)$ | $P_6(y==1)$ |
| $P_7(y==0)$ | $P_7(y==1)$ |
| $P_8(y==0)$ | $P_8(y==1)$ |
| $P_9(y==0)$ | $P_9(y==1)$ |
| $P_{10}(y==0)$ | $P_{10}(y==1)$ |

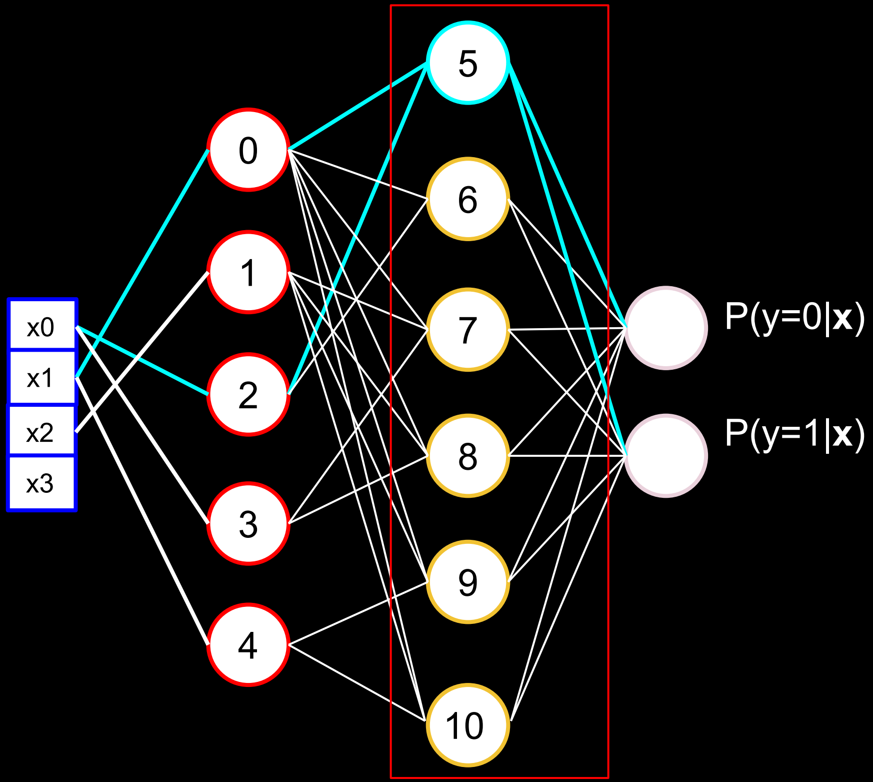
Select the corresponding row

**result = op1W2**

**x** =

| | | | |
|---|---|---|---|
| 0.1 | 0.8 | 0.2 | 0.1 |



# Decision Tree in 5 Lines of Python

```
op0 = np.dot(x, W0) + b0
op0 = 2*(op0 >= 0) - 1  # non linearity

op1 = np.dot(op0, W1) + b1
op1 = op1 >= 0

op_probability = np.dot(op1, W2)  # result
```

With no conditional statements!

# RF as ANN

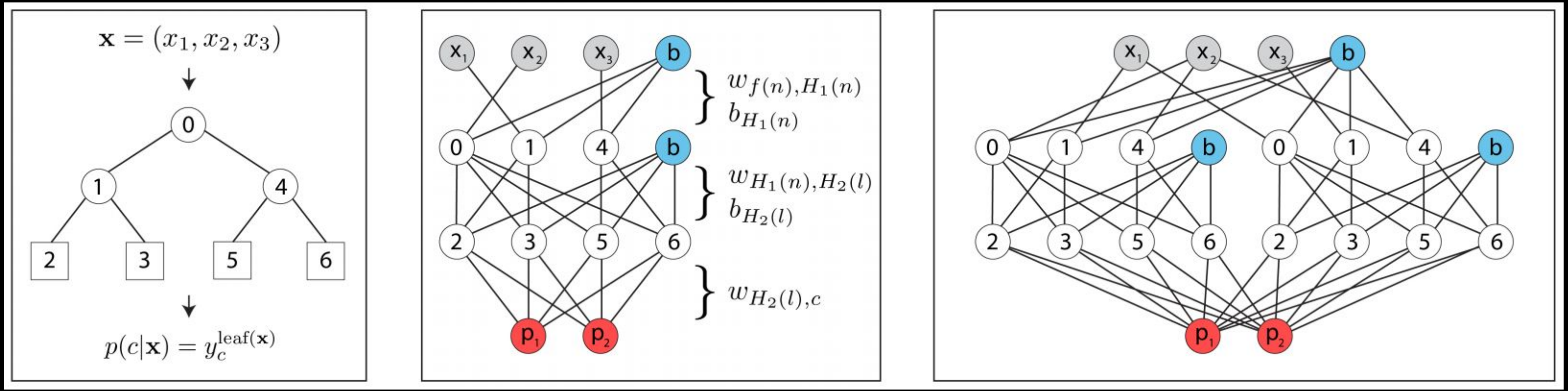



Fig from Richmond 2015

# Experiments

| | RF | ANN | Sparse | Relaxed | Vote |
|---|---|---|---|---|---|
| Wisconsin | 4.8 (1.7) | 3.4 (1.7) | **2.9 (1.3)** | 3.5 (1.9) | 4.1 (1.8) |
| Ionosphere | 6.5 (2.2) | 11.0 (3.3) | **6.2 (2.0)** | 6.9 (2.7) | 6.7 (1.9) |
| Sonar | 21.5 (5.1) | 18.0 (5.5) | **14.4 (4.8)** | 16.0 (5.2) | 14.8 (4.1) |
| Landsat | 10.9 (0.3) | 10.1 (0.4) | 9.1 (0.3) | 9.1 (0.4) | **9.0 (0.4)** |
| Pima | **24.5 (2.1)** | 28.9 (3.0) | 26.8 (3.1) | 26.4 (2.3) | 26.6 (2.6) |
| Heart | **16.3 (4.4)** | 21.8 (4.3) | 19.5 (4.1) | 19.5 (4.3) | 19.0 (4.1) |
| Credit | 25.6 (2.2) | 28.3 (2.7) | **24.6 (1.8)** | 24.7 (2.9) | 25.3 (2.7) |

# Cascaded RF as CNN

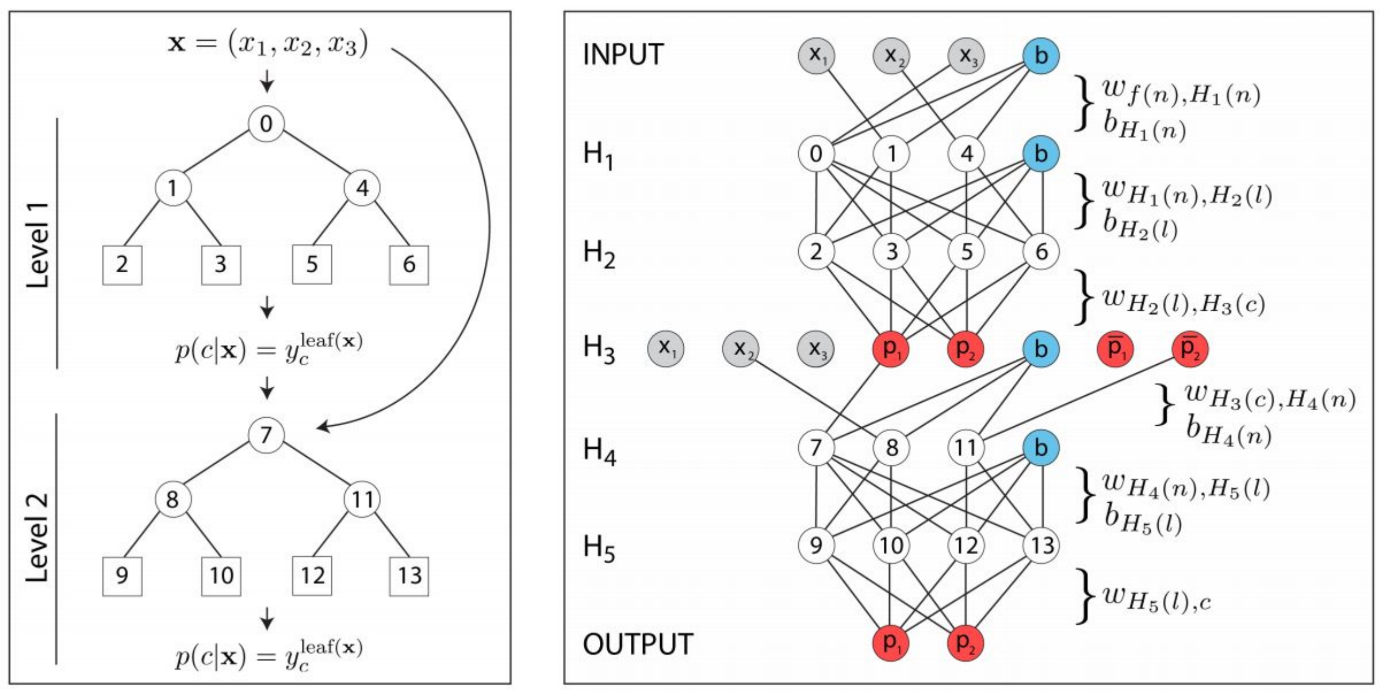



Fig from Richmond 2015