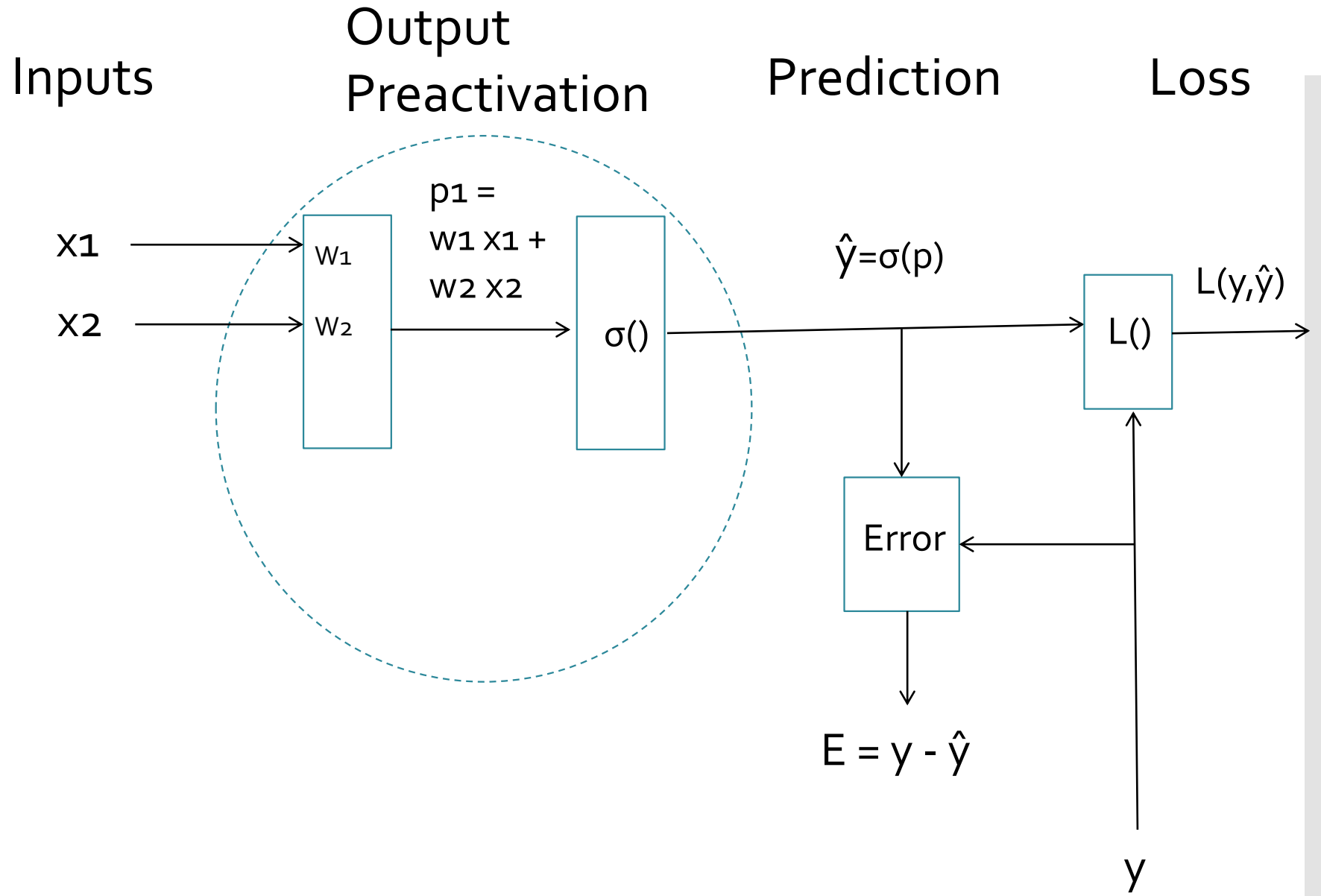


# Cascor: Three Perspectives on Error

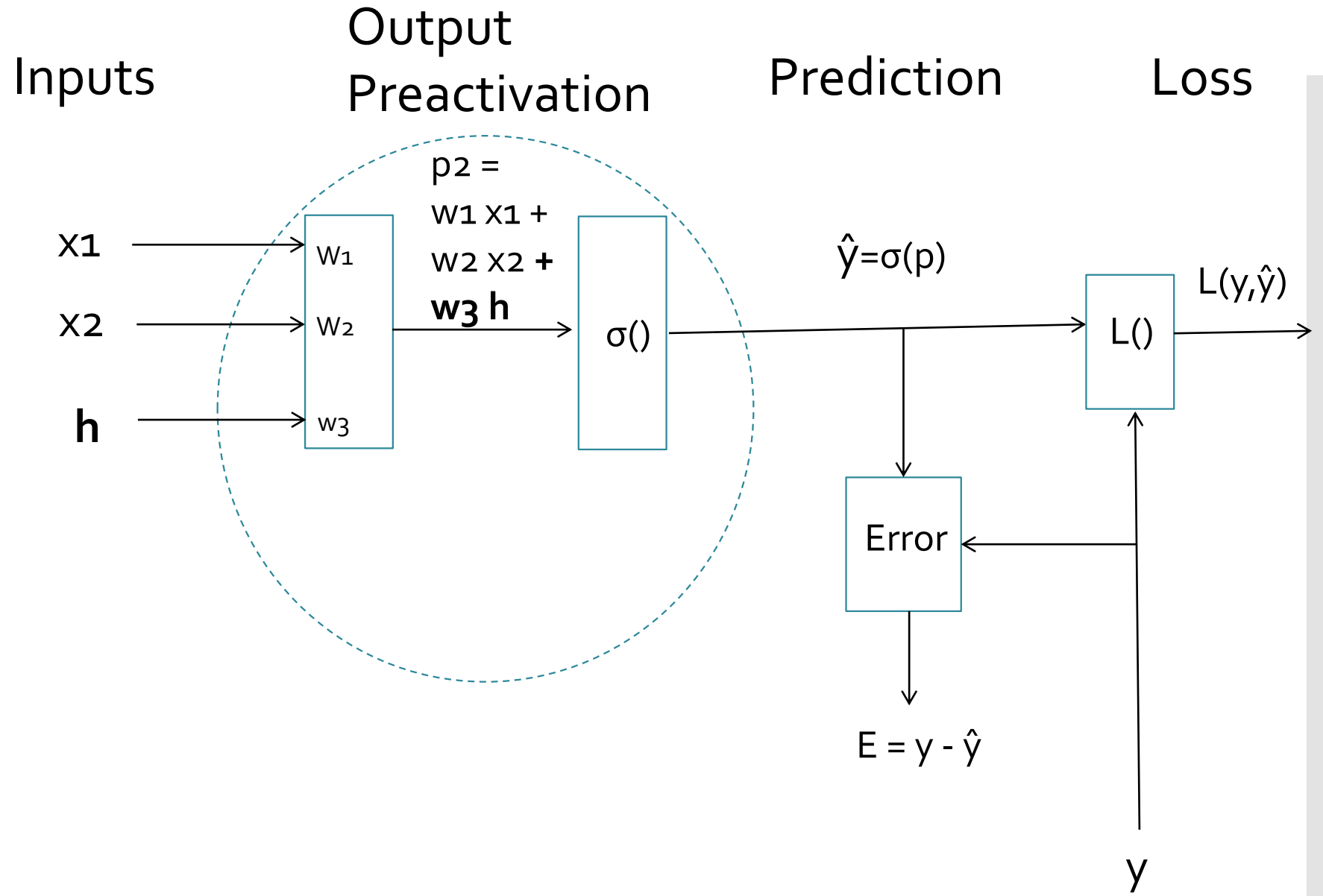
# Simplest Network



## Cascor Goal

- **Cascor Training Goal**
  - Add a hidden unit that reduces training error
- **Creating a Hidden Unit**
  - Train a hidden unit to correlate with error
  - Error is *information* on how to predict better

# Adding a Hidden Unit



# Example

- **Cascor for classification**
  - Output activation function:
    - logistic (sigmoid)

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

- Loss function:
  - binary cross entropy

$$L(y_i, \hat{y}_i) = - \sum y_i \sigma(\hat{y}_i) + (1 - y_i)(1 - \sigma(\hat{y}_i))$$

# Example: Two Data Points

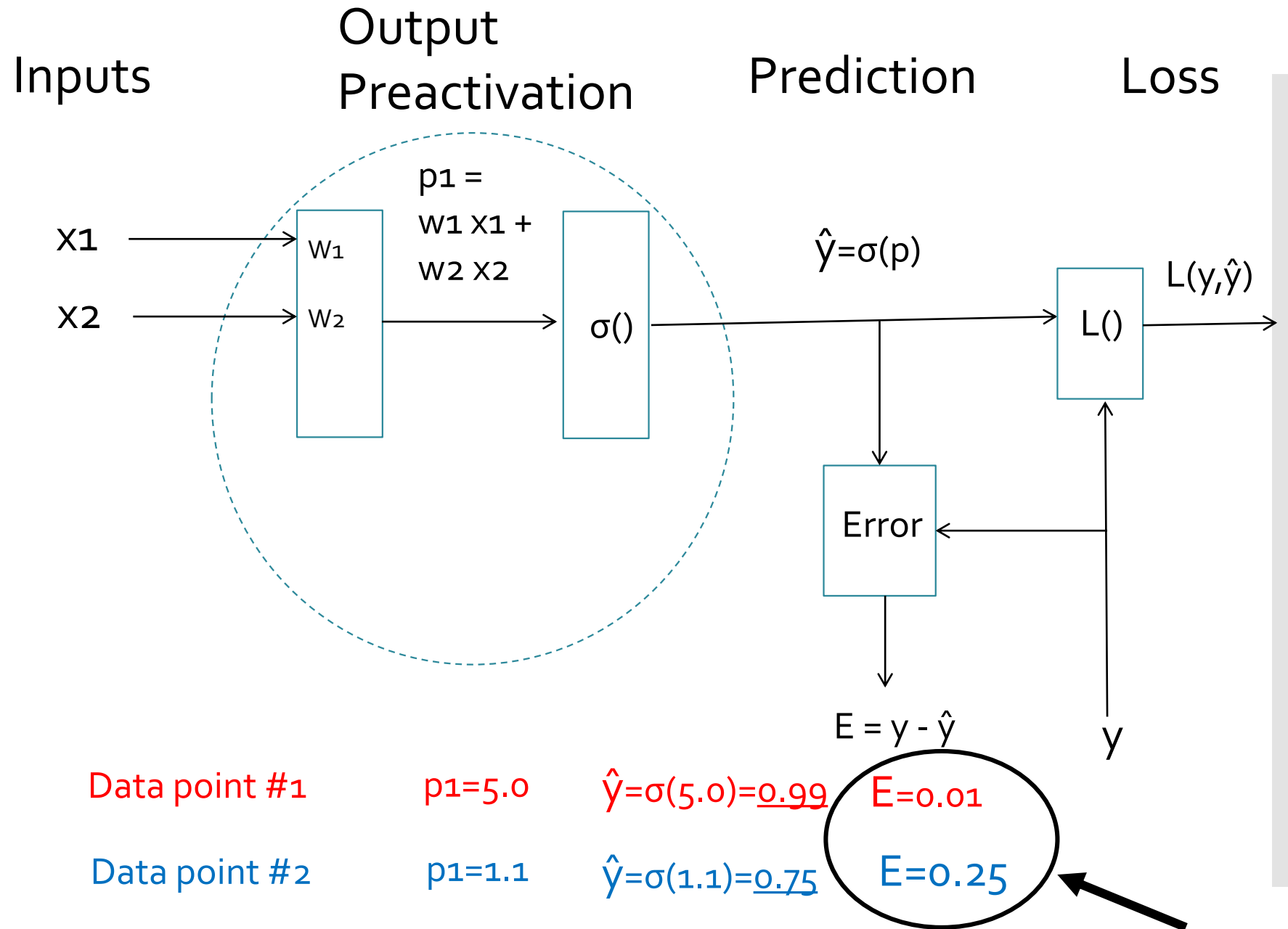
Data Point	Output Preactivation	Output Prediction	Label	Loss
#1	$p_1=5.0$	$\sigma(5.0)=0.99$	1	$-\ln(0.99)=0.007$
#2	$p_1=1.1$	$\sigma(1.1)=0.75$	1	$-\ln(0.75)=0.29$

This assumes no hidden unit  
added yet

# Error

- **Three Perspectives on Error**
  1. (Traditional) “My prediction was bad”

Error

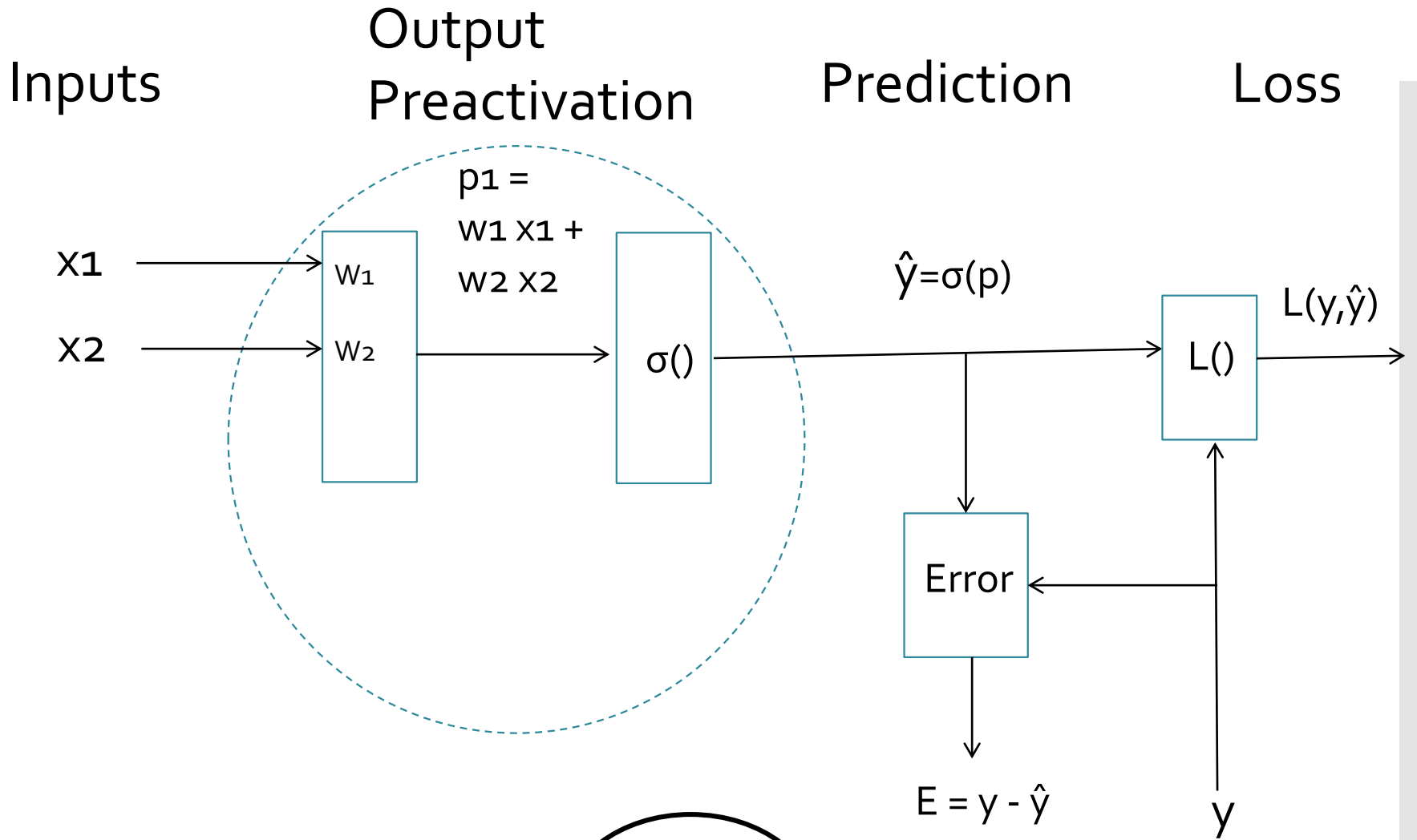
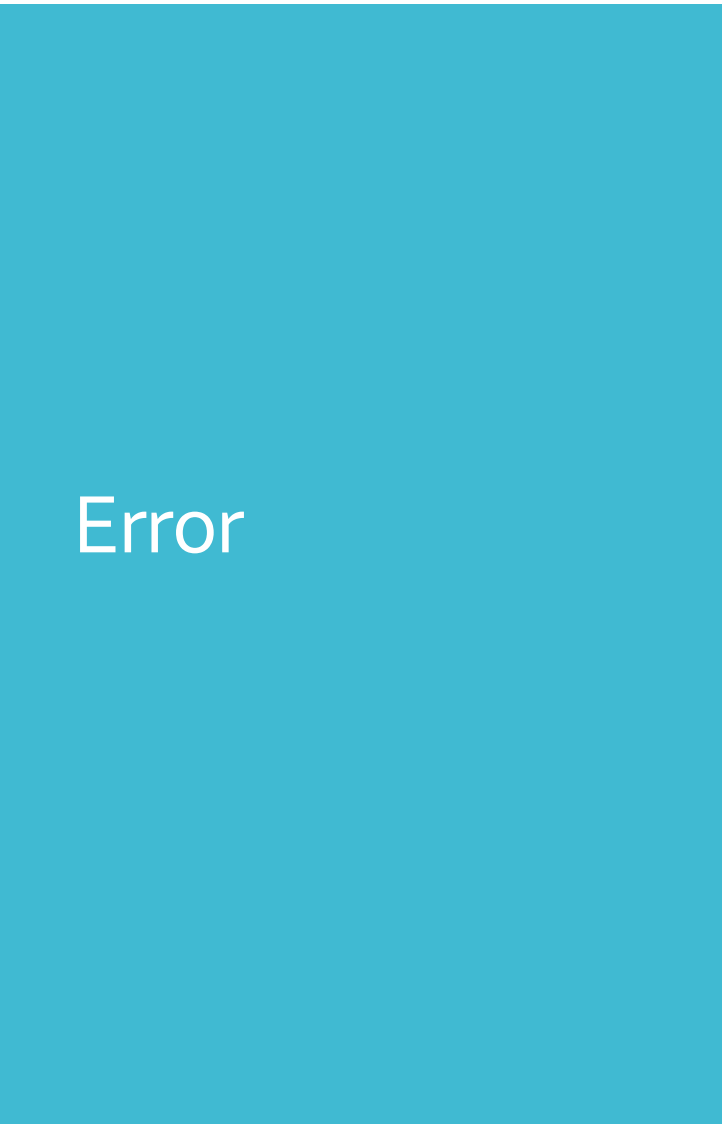




# Error

- **Three Perspectives on Error**

1. (Traditional) “My prediction was bad”
2. “My preactivation was bad”



Data point #1

$p_1 = 5.0$

$\check{E} \approx 0$

$\hat{y} = \sigma(5.0) = 0.99$

Data point #2

$p_1 = 1.1$

$\check{E} = 3.9$

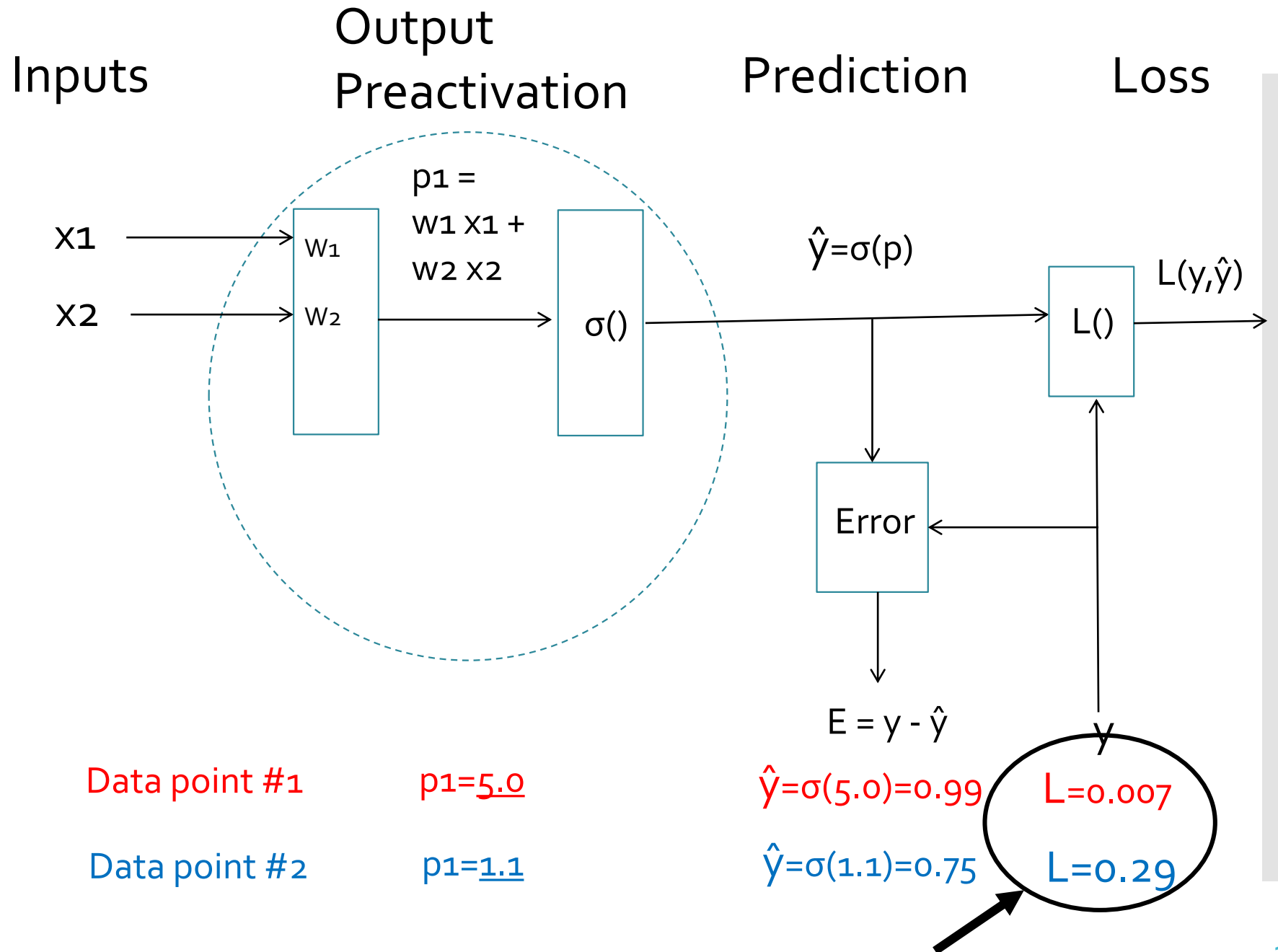
$\hat{y} = \sigma(1.1) = 0.75$

# Error

- **Three Perspectives on Error**

1. (Traditional) “My prediction was bad”
2. “My preactivation was bad”
3. “My loss was bad”

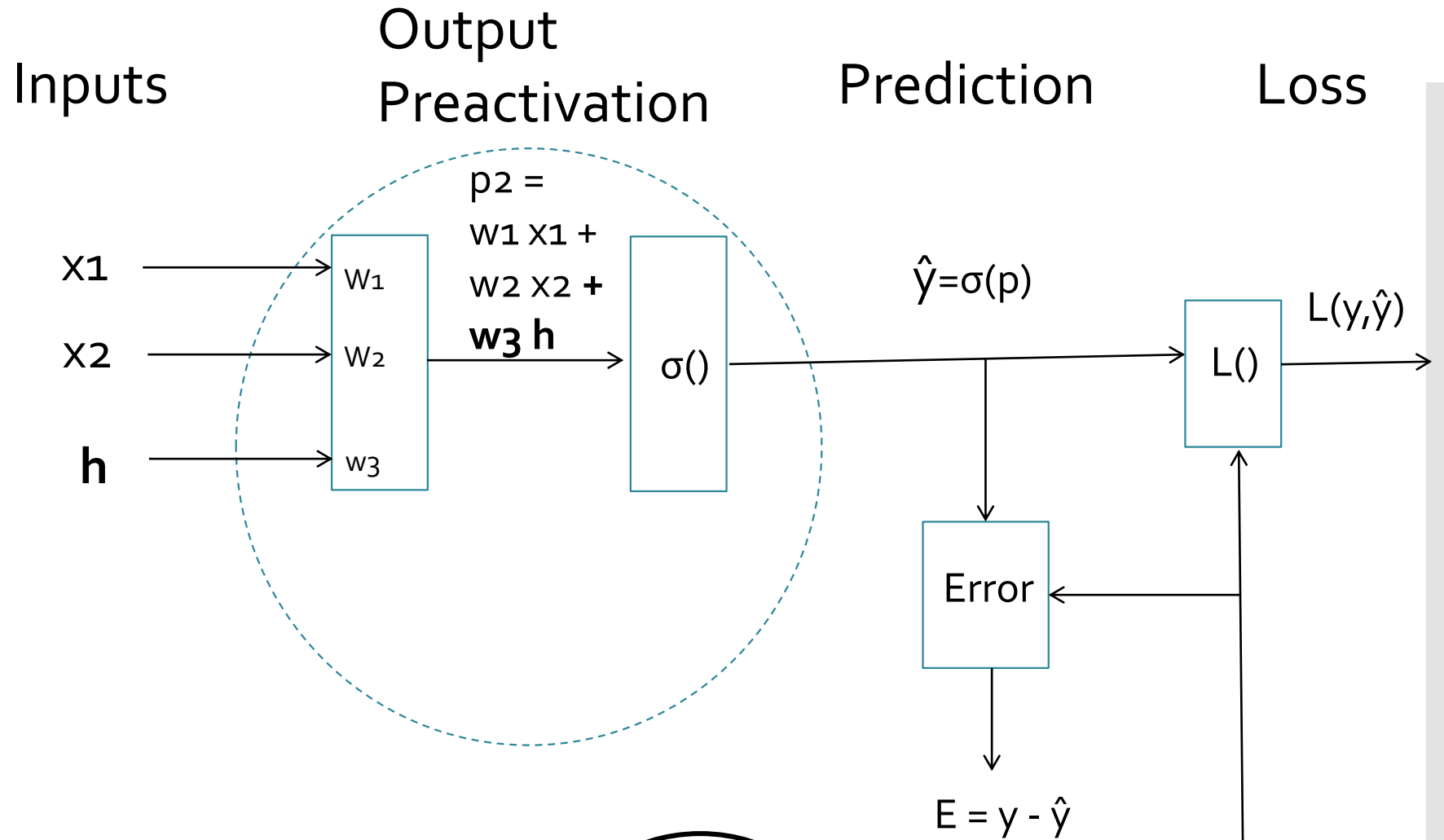
Error



## Error

- **Three Perspectives on Error**
  1. (Traditional) “My prediction was bad”
  2. “My preactivation was bad”
  3. “My loss was bad”
- **Can train hidden unit to correlate with *any* kind of error**
  - But maybe #2 makes the most sense
    - (since hidden unit adds to preactivation)

# Adding a Hidden Unit



Data point #1

$p_1 = 5.0$   $\check{E} \approx 0$

$\hat{y} = \sigma(5.0) = 0.99$

Data point #2

$p_1 = 1.1$   $\check{E} = 3.9$

$\hat{y} = \sigma(1.1) = 0.75$

What if  $h$  adds this amount?

## Preactivation Error

- **If we train a hidden unit to correlate with the *error in preactivation***
  - the hidden unit will add an amount close to the 'missing' preactivation

## Practivation Error

- **If we train a hidden unit to correlate with the *error in preactivation***
  - the hidden unit will add an amount close to the 'missing' preactivation
- **BUT**
  - Loss function is nonlinear
  - Some incremental amounts of preactivation better for loss than others
  - **Correlate with preactivation's incremental loss improvement?**