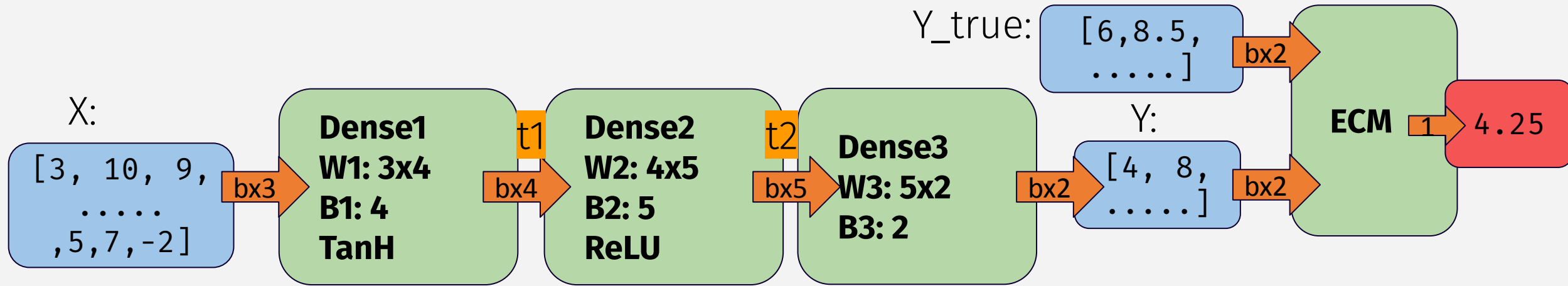


# Uso de memoria de una red

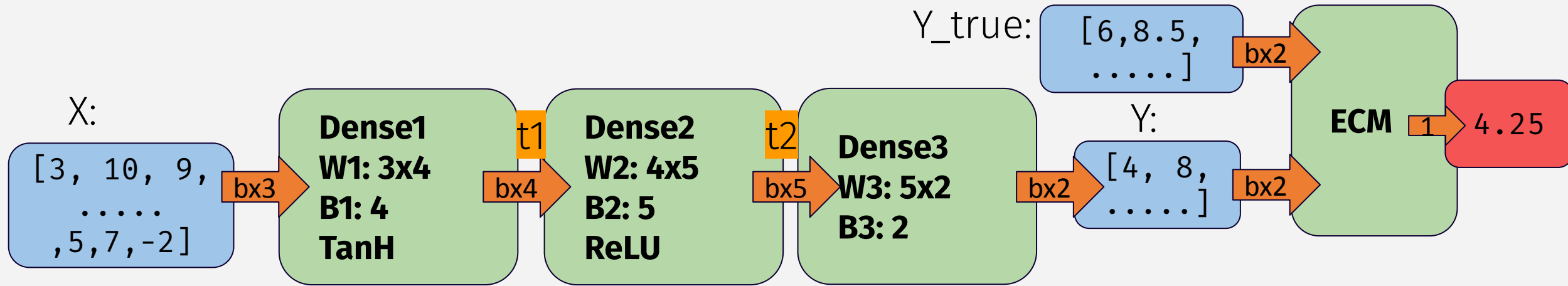
---

# Uso de memoria de una red



- **Base Estática**
  - Parámetros de la red
- **Forward** (b ejemplos)
  - Salidas de cada capa
  - Valores necesarios para el backward
- **Backward** (b ejemplos)
  - Derivadas de parámetros
  - Derivadas de entradas
  - Valores intermedio

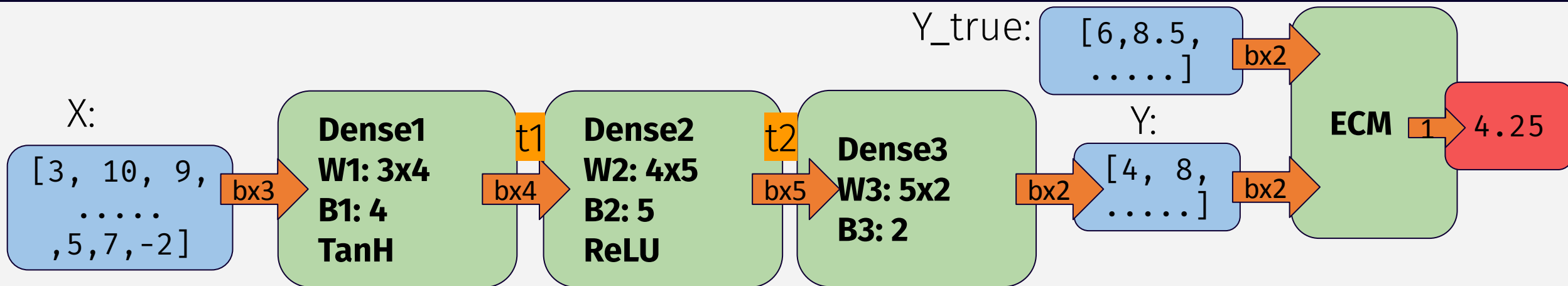
# Uso de memoria base o estática



- Dense1: 16 valores
  - $W1: 12$
  - $B1: 4$
- Dense2: 25 valores
  - $W2: 20$
  - $B2: 5$
- Dense3: 12 valores
  - $W2: 10$
  - $B3: 2$

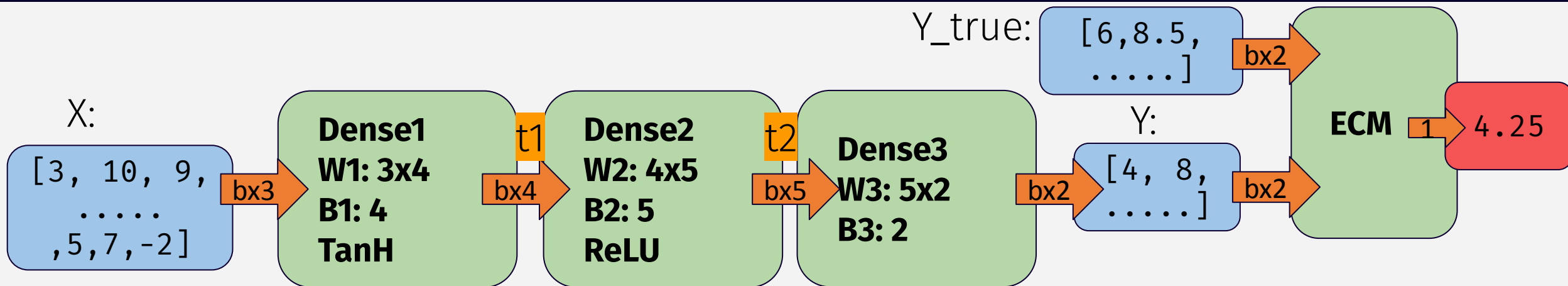
- Total = 53 valores
  - $16+25+12$
- Funciones de activación
  - No tienen parámetros
- Error
  - No tiene parámetros

# Forward con lote de b ejemplos



- Entradas
  - $x: b \times 3$
  - $y_{true}: b \times 2$
- Valores intermedios
  - $t1: b \times 4$
  - $t2: b \times 5$
  - $Y: b \times 2$
  - $E = 1$
- Total:  $b * (3+2+4+5+2) + 1 = b * 17$
- Total depende de  $b$ 
  - Ajustar  $b$  respecto a memoria disponible
- Puede requerir más memoria
  - Guardar valores p/ backward
  - Cálculos intermedios
- Puede requerir menos memoria
  - Desechar valores intermedios
  - $\max(b \times 4, b \times 5, b \times 2, 1) = b \times 5$

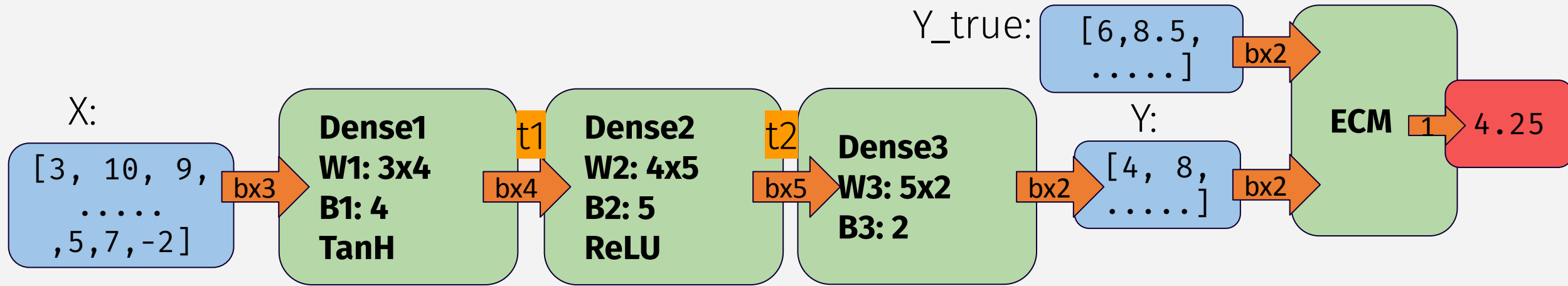
# Backward con lote de b ejemplos



- Gradientes de parámetros
  - Descenso de Gradiente básico
    - 1 parámetro  $\rightarrow$  1 gradiente
      - 53 valores
  - Descenso de gradiente avanzado (ej: con inercia)
    - 1 parámetro  $\rightarrow$  1 gradiente + 1 inercia
      - 106 valores

- Gradientes de intermedios
  - 1 intermedio  $\rightarrow$  1 gradiente
  - $b * 17$  valores
- Puede requerir más memoria
  - Cálculos intermedios
- Puede requerir menos memoria
  - Desechar valores intermedios
  - $\max(b*4, b*5, b*2, 1) = b*5$

# Resumen de uso de memoria



- Red
  - $x: b \times 3$
  - $y\_true: b \times 2$
  - Dense1/Dense2/Dense3/ECM
- Estático: 53 valores
- Forward:
  - $b \times 17$  valores (intermedios)
- Backward
  - gradientes + intermedios
  - $53 + b \times 17$  valores
- No dependen de optimizaciones
  - Estático
  - Gradientes
- Depende de optimizaciones
  - Intermedios
  - En la práctica, difícil desechar intermedios
    - Reasignar memoria es caro
    - Difícil es grafos no secuenciales