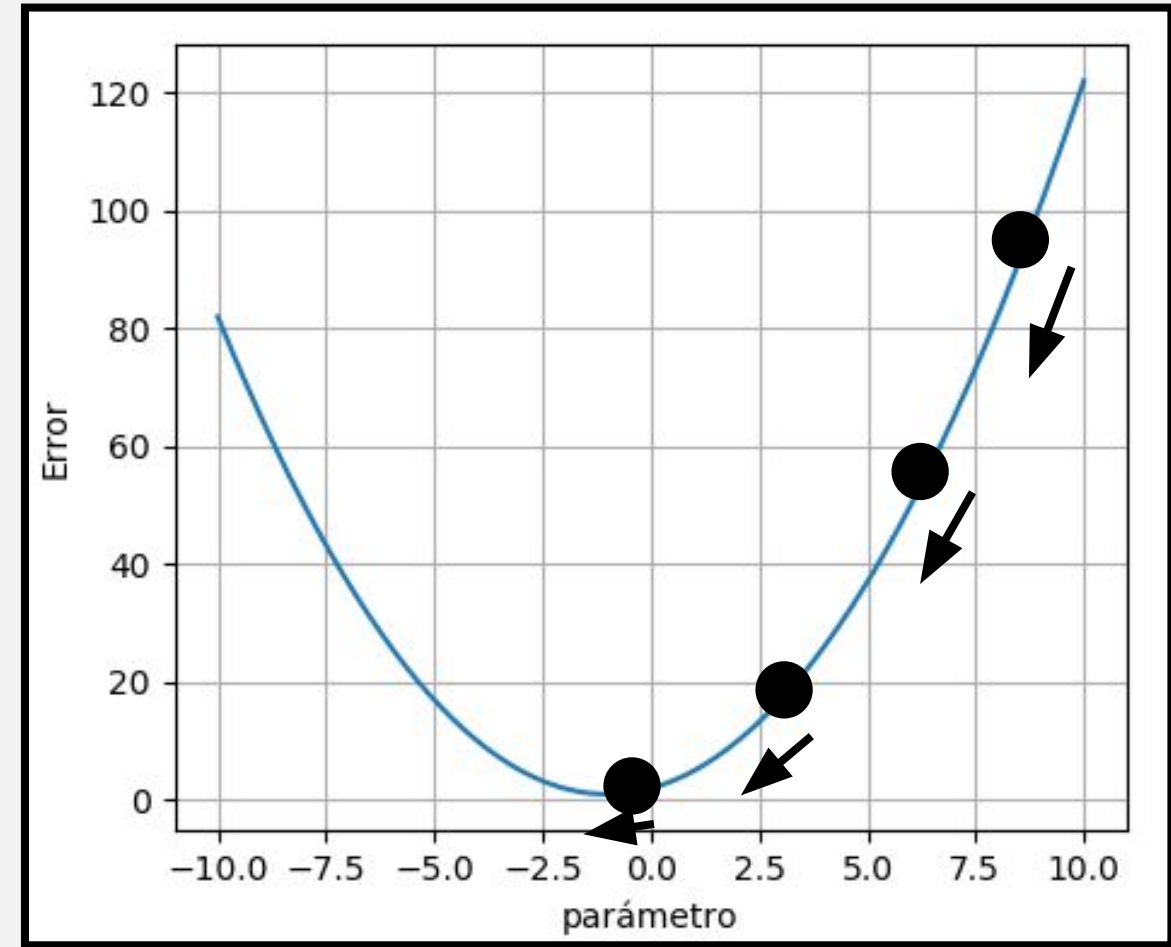


# Monitoreo y Terminación de Descenso de Gradiente

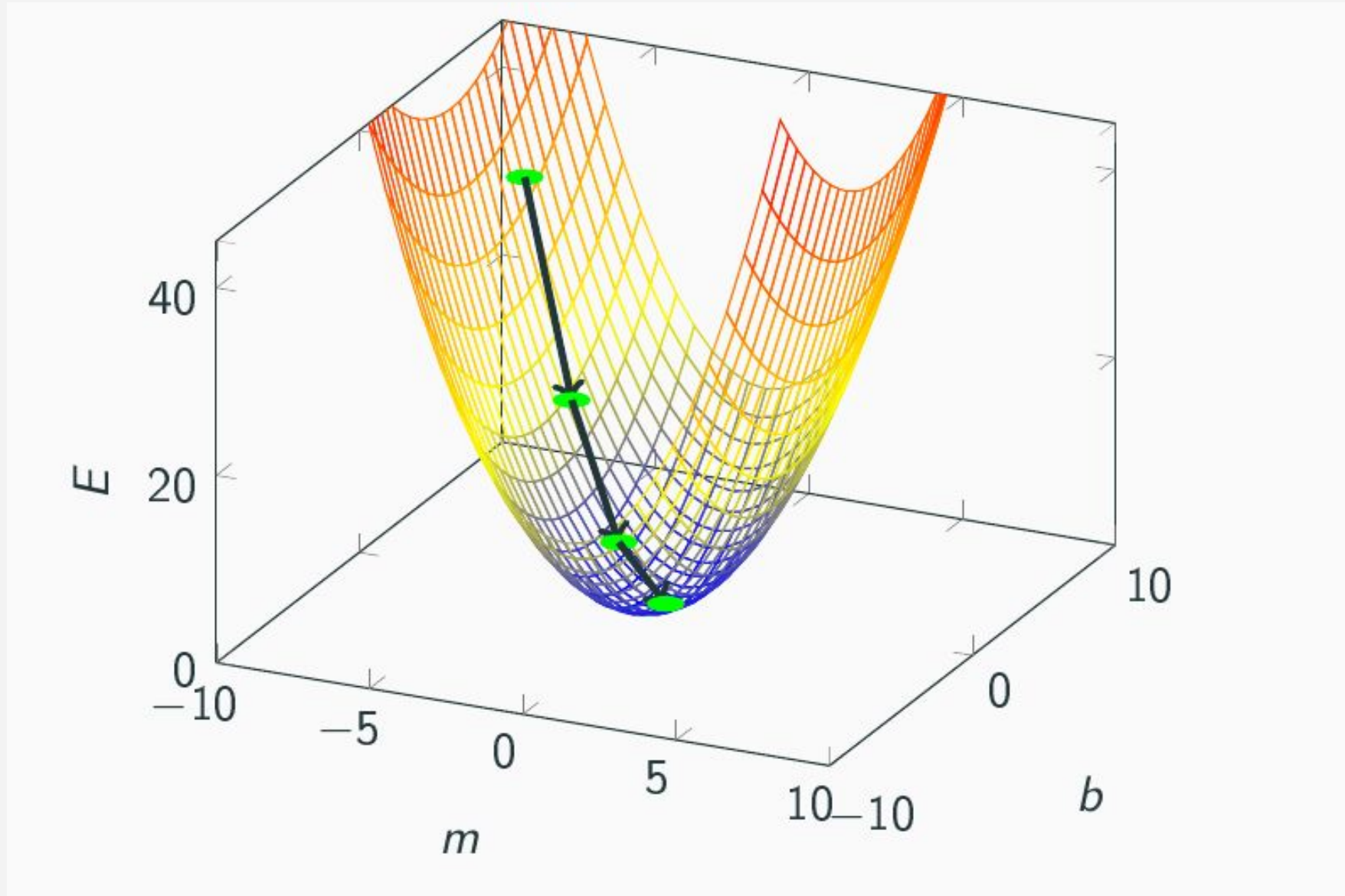
---

# Descenso de gradiente

- Descenso de gradiente
  - Iterativo
  - Generalizable
    - Regresión Lineal
    - Regresión Logística
    - Redes Neuronales
    - Máquinas de Vectores de Soporte
      - Cualquier modelo con  $E$  derivable
  - Escalable (con modificaciones)
    - Millones de ejemplos



# Descenso de gradiente con 2 parámetros



# Descenso de gradiente con $n > 2$ parámetros

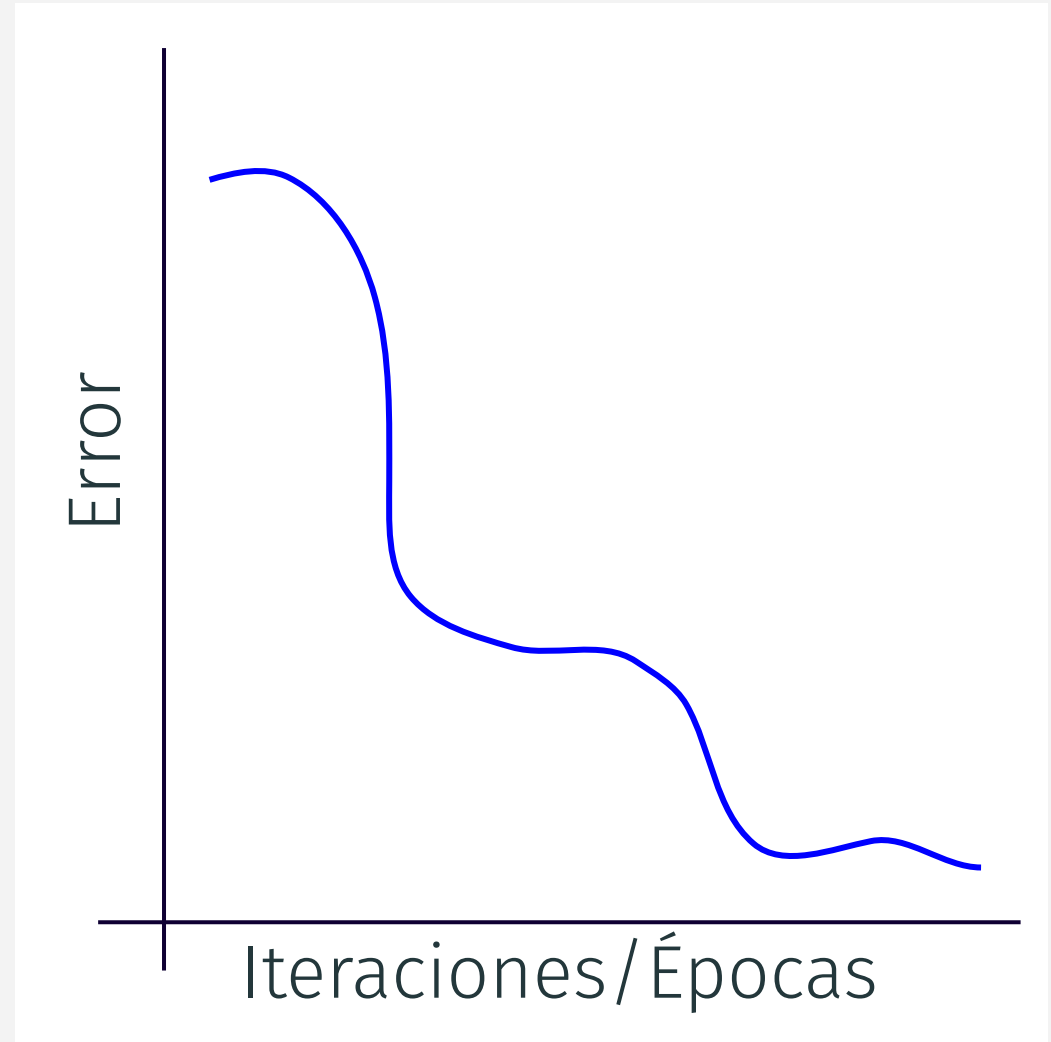
- ¿Cuándo terminar?
  - Iteraciones fijas
  - Tolerancia de error
- ¿Cómo determinar?
  - Monitorear parámetros vs error
    - Ver si se llega al mínimo
    - No se puede visualizar E con  $n > 2$  parám.



Error vs Parámetros

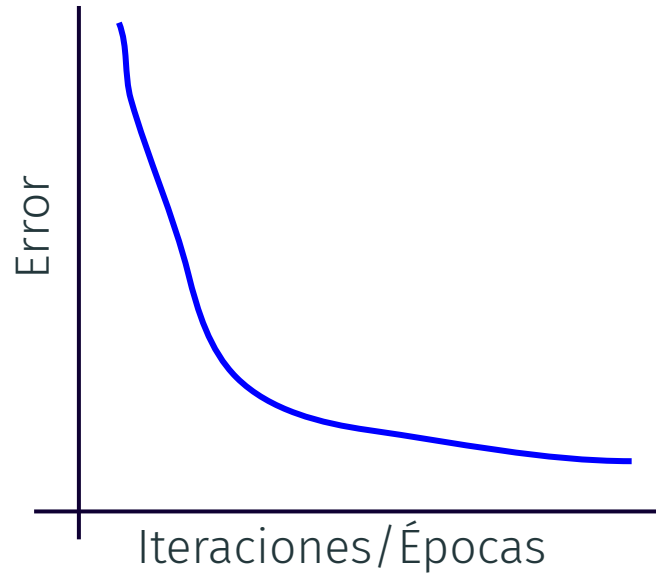
# Curvas de error

- Monitorear Error vs Iteración
  - No requiere 1D o 2D
  - Funciona si el error no es convexo
- Cada Iteración
  - Calcular error
  - Actualizar gráfico
  - (o guardar en vector y dibujar al final)
  - (idem en un archivo)
- Sirve para otros métodos iterativos

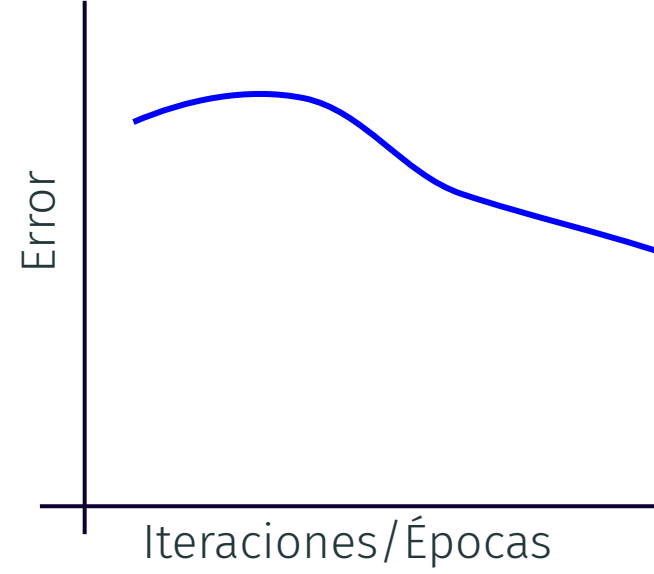
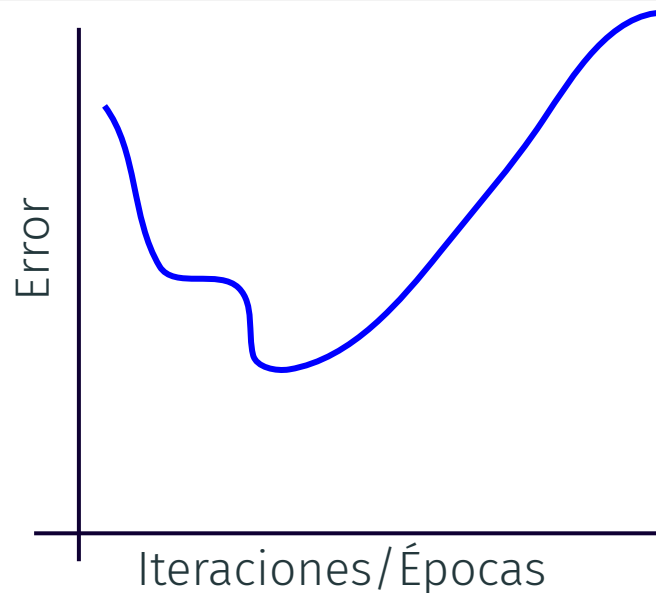


# Curvas de error - Posibles síntomas

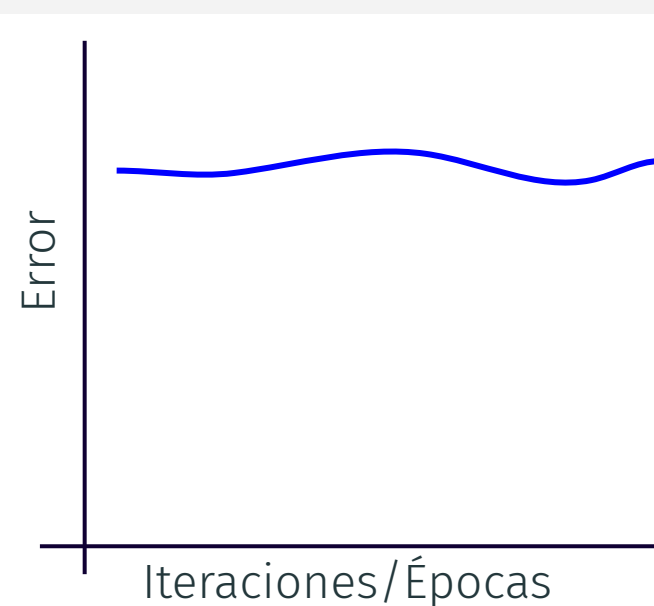
- Convergencia Rápida
- Ajustes menores
- Típico



- Divergencia
- Posibilidades
- 1)  $\alpha$  muy alto
- 2) ej. sin normalizar



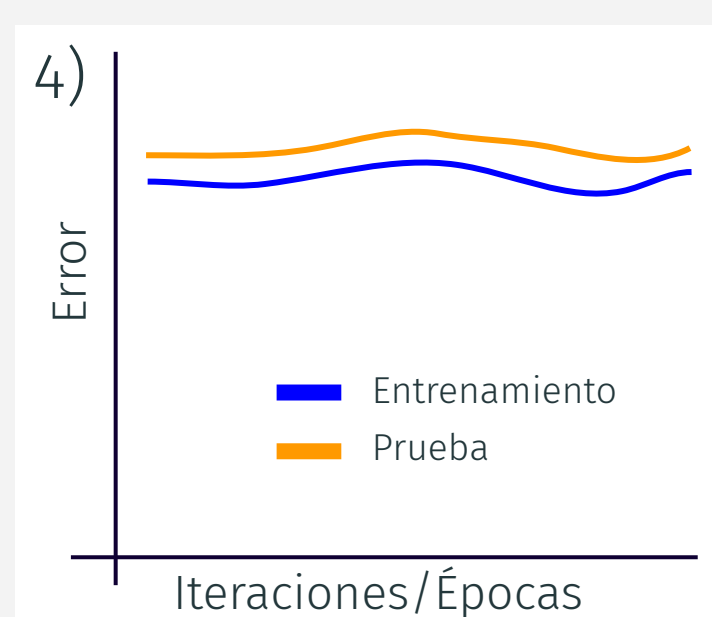
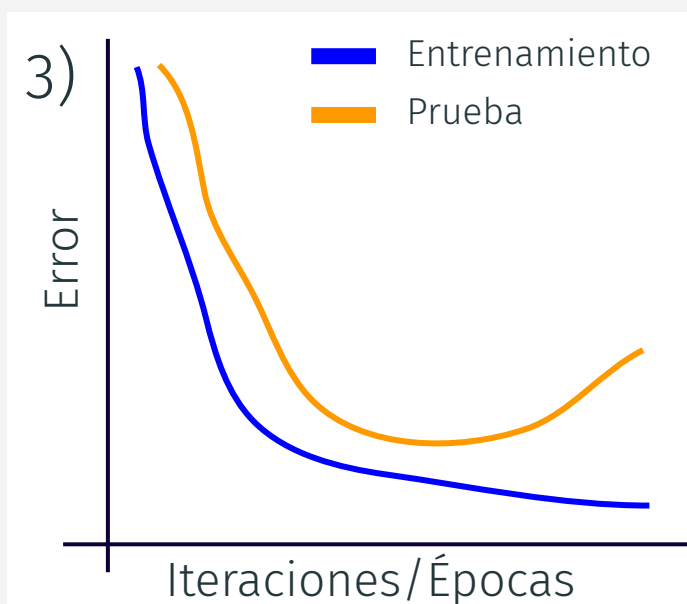
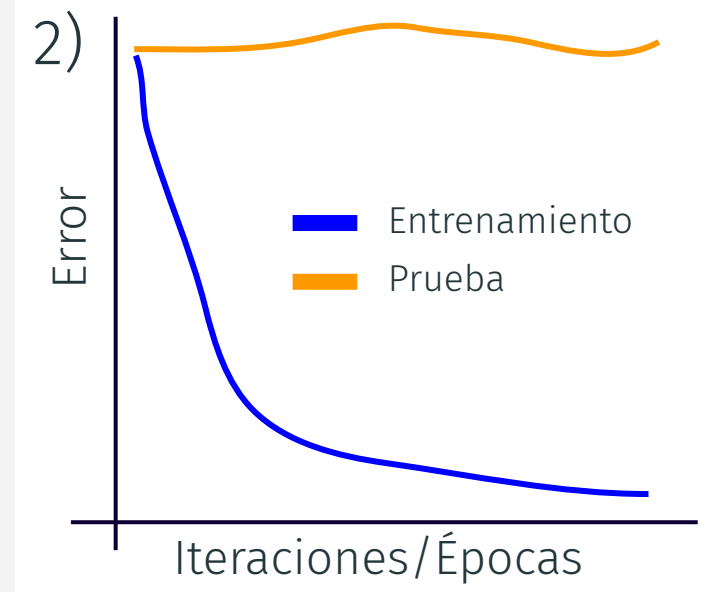
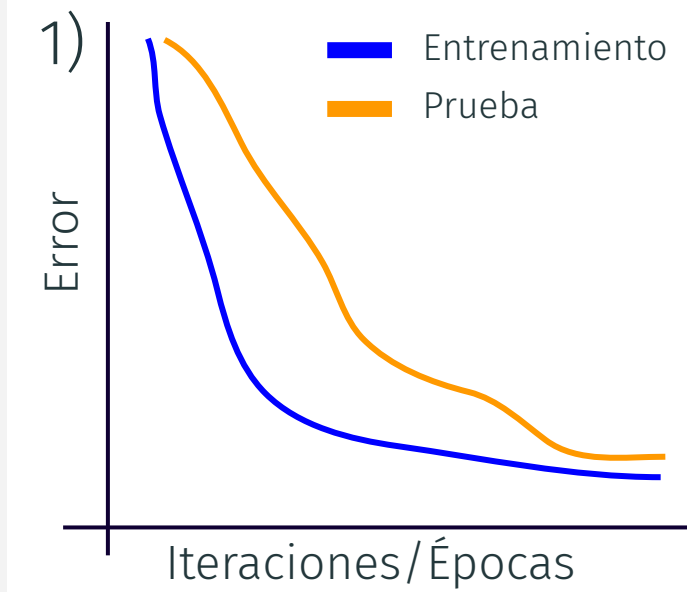
- Ajustes constantes
- Posibilidades
- 1) Faltan iteraciones
- 2)  $\alpha$  muy chico



- Modelo no aprende
- Problema muy complejo
- $\alpha$  muy chico

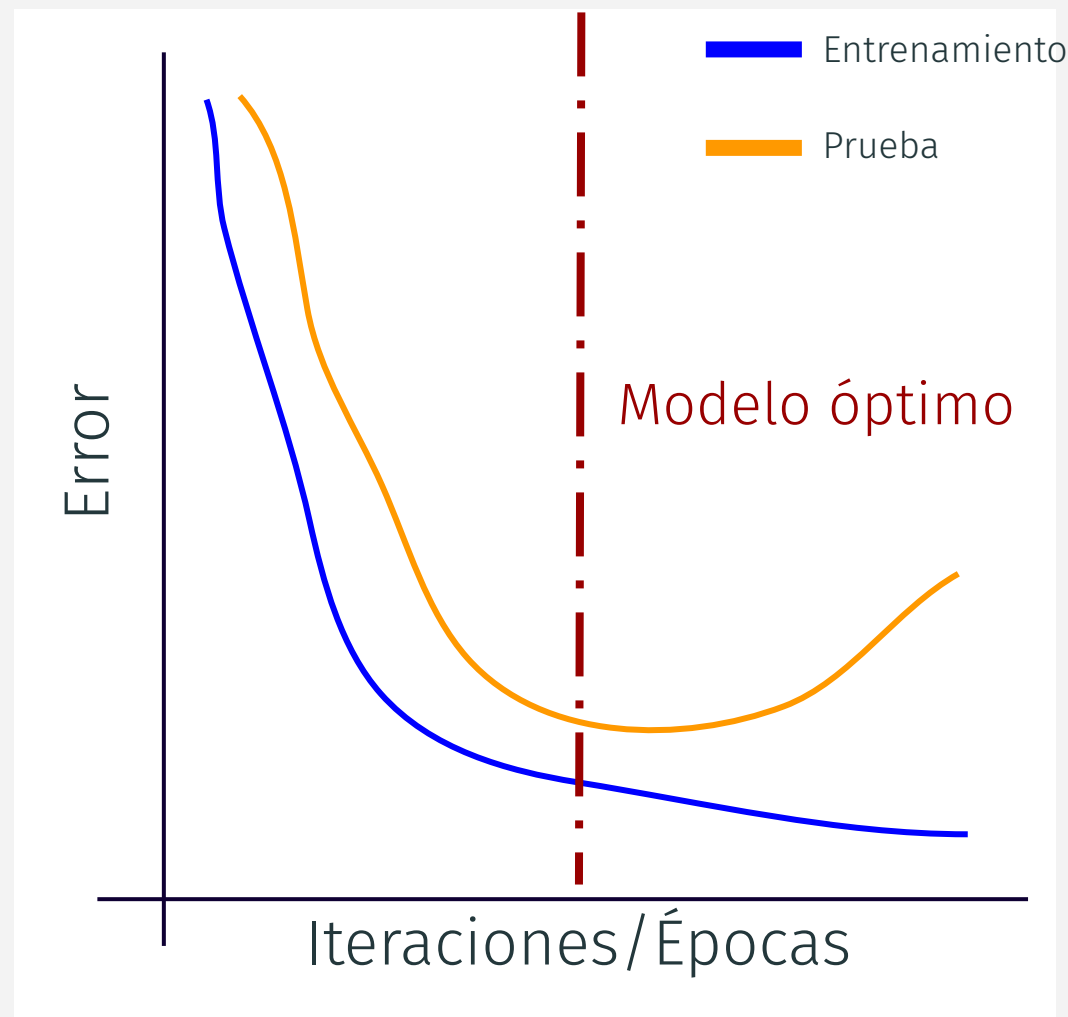
# Curvas de error entrenamiento/prueba

- 1)
  - Buen entrenamiento
  - Buena generalización
- 2)
  - Buen entrenamiento
  - Mala generalización
- 3)
  - Buen entrenamiento
  - Buena generalización
  - Mala a lo último
    - Parar antes
- 4)
  - Mal entrenamiento
  - Mala generalización
  - Cambiar modelo



# Para antes (early stopping)

- Entrenar más tiempo
  - Aprender mejor
    - Conjunto entrenamiento
- Típico
  - K iteraciones
    - Error de prueba ↓
  - K+1 en adelante
    - Error de prueba ↑
- Early Stopping
  - No esperar a que error en entrenamiento se estanque
  - Monitorear error en prueba (o val)



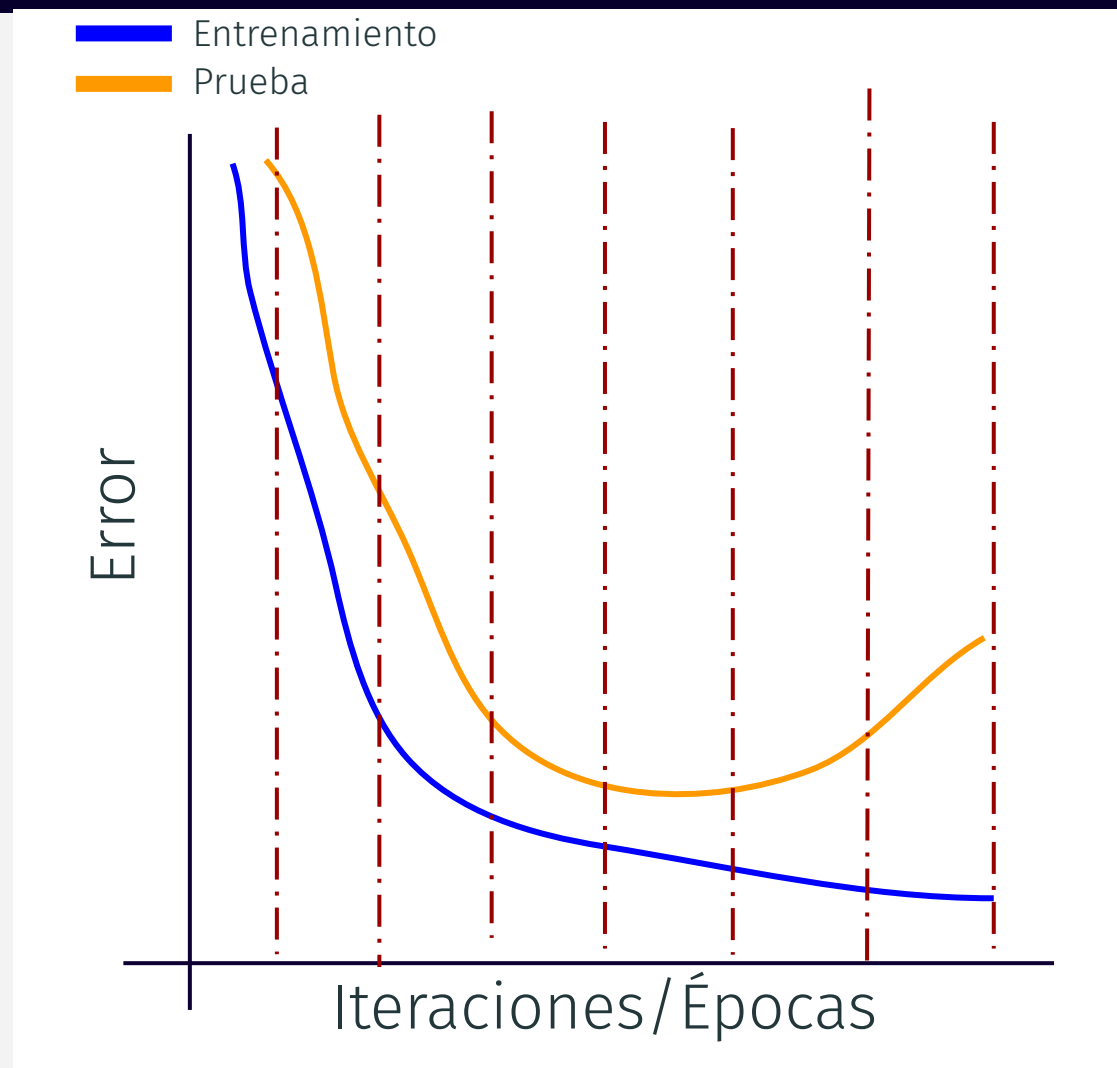


# Descenso de gradiente con early stopping

- **descenso\_gradiente**( $x_{\text{train}}, y_{\text{train}}, x_{\text{test}}, y_{\text{test}}, \alpha, \text{ite}$ )
  - $p$  = parametros aleatorios
  - $E = \text{error}(x_{\text{test}}, y_{\text{test}}, p)$
  - $i = 0$
  - Repetir
    - Calcular  $\delta E / \delta p$
    - $p = p - \alpha \delta E / \delta p$
    - $E_{\text{anterior}} = E$
    - $E = \text{error}(x_{\text{test}}, y_{\text{test}}, p)$
    - $i = i + 1$
    - Hasta que  $\text{abs}(E - E_{\text{anterior}}) \simeq 0$  o  $i = \text{ite}$
  - retornar  $p$

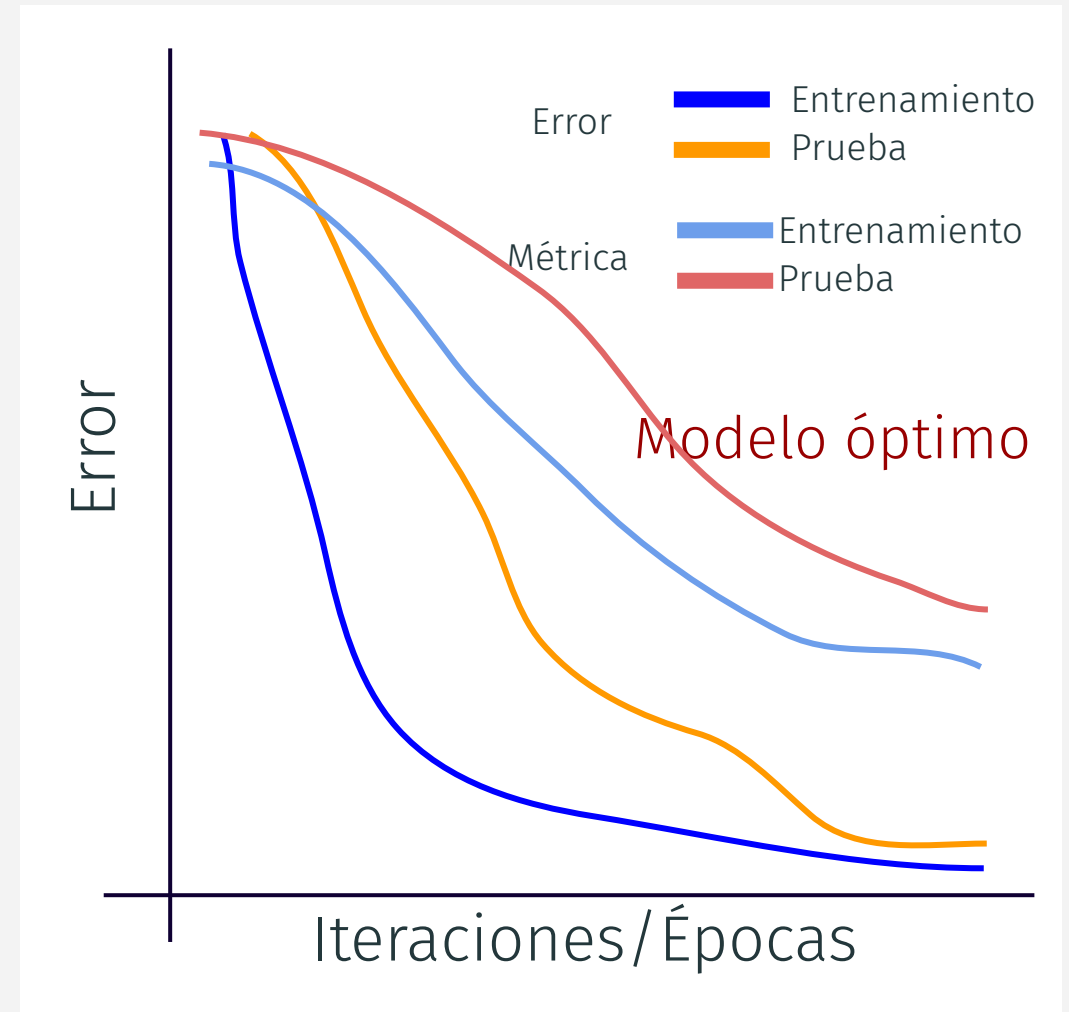
# Para antes (early stopping)

- Alternativamente
  - Entrenar hasta el final
  - Guardar *checkpoints* del modelo
    - Valores de los parámetros
    - En distintas iteraciones
  - Elegir el *checkpoint* con menor error de prueba



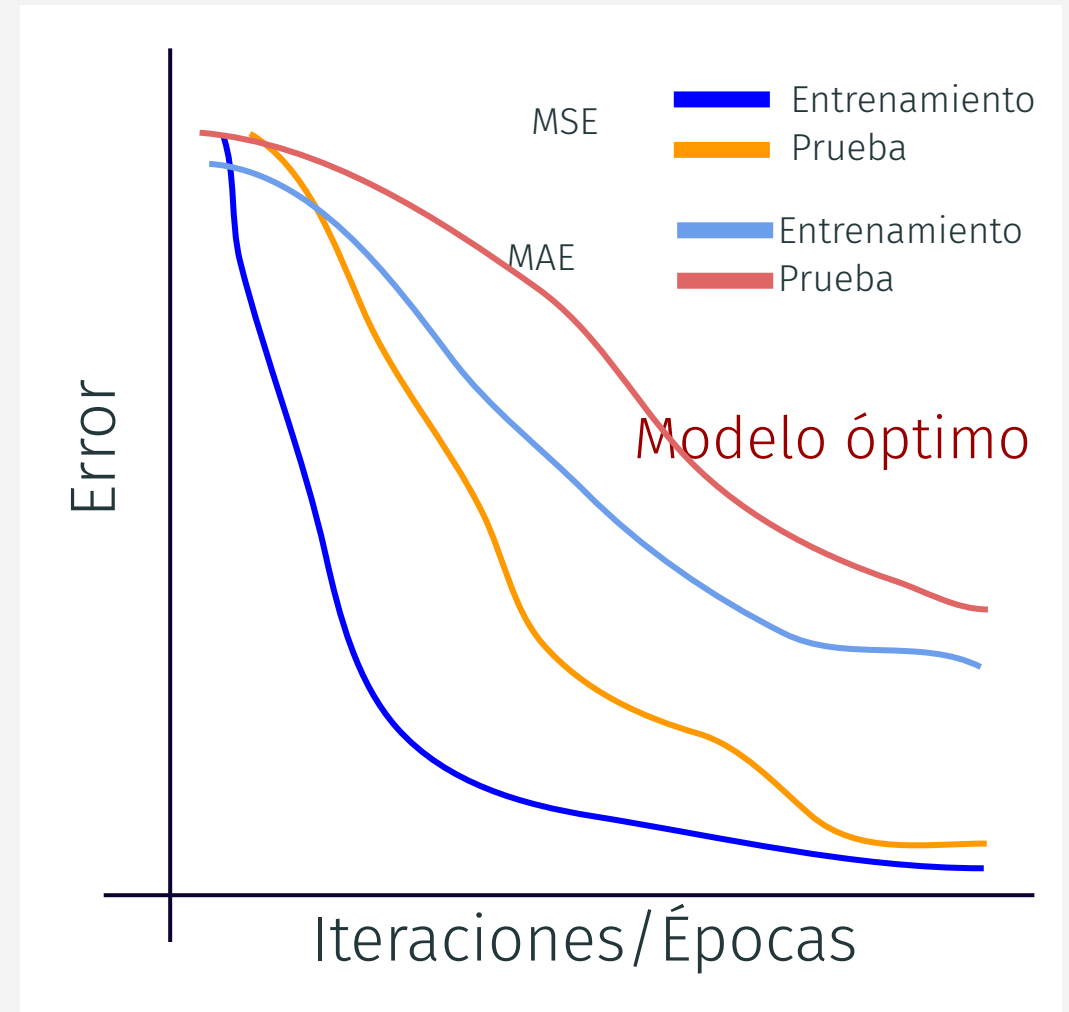
# Métricas de error

- Función de error E
  - Diseñada para optimizar
  - Ejemplo MSE
    - $MSE_i = (y_i - f(x_i))^2$
    - $MSE = (1/n) \sum_i^n MSE_i$
    - Difícil de interpretar  $MSE_i$ 
      - Difícil de interpretar MSE
    - Si **MSE = 12** ¿qué significa?
- Métricas
  - Similares a funciones de error
  - No tienen que ser optimizables
  - Más fáciles de interpretar
  - Fáciles de comparar entre modelos / problemas



# Métrica Error Absoluto Medio

- Métrica MAE
  - Mean Absolute Error
  - Para problemas de Regresión
  - $MAE_i = \text{abs}(y_i - f(x_i))$
  - $MAE = (1/n) \sum_i^n MAE_i$
  - No es derivable
    - $\text{abs}()$  no es derivable en 0
- Fácil de interpretar
  - Mismas unidades que la salida **y**



# Ejemplo: Error Absoluto Medio

- Métrica MAE
  - $MAE_i = \text{abs}(y_i - f(x_i))$
  - $MAE = (1/n) \sum_i^n MAE_i$
- Error MSE
  - $MSE_i = (y_i - f(x_i))^2$
  - $MSE = (1/n) \sum_i^n MSE_i$

Salida	Esperado	norm <sup>2</sup>	MSE	abs(norm)	MAE
3	5	4	<b>10</b>	2	<b>3</b>
3	5	4		2	
1	5	16		4	
3	-1	16		4	

MAE: En promedio, el modelo tiene un error de 3 puntos