

Trabajo Práctico Especial 2

Redes Neuronales

Q1-2014 ITBA

Grupo 5

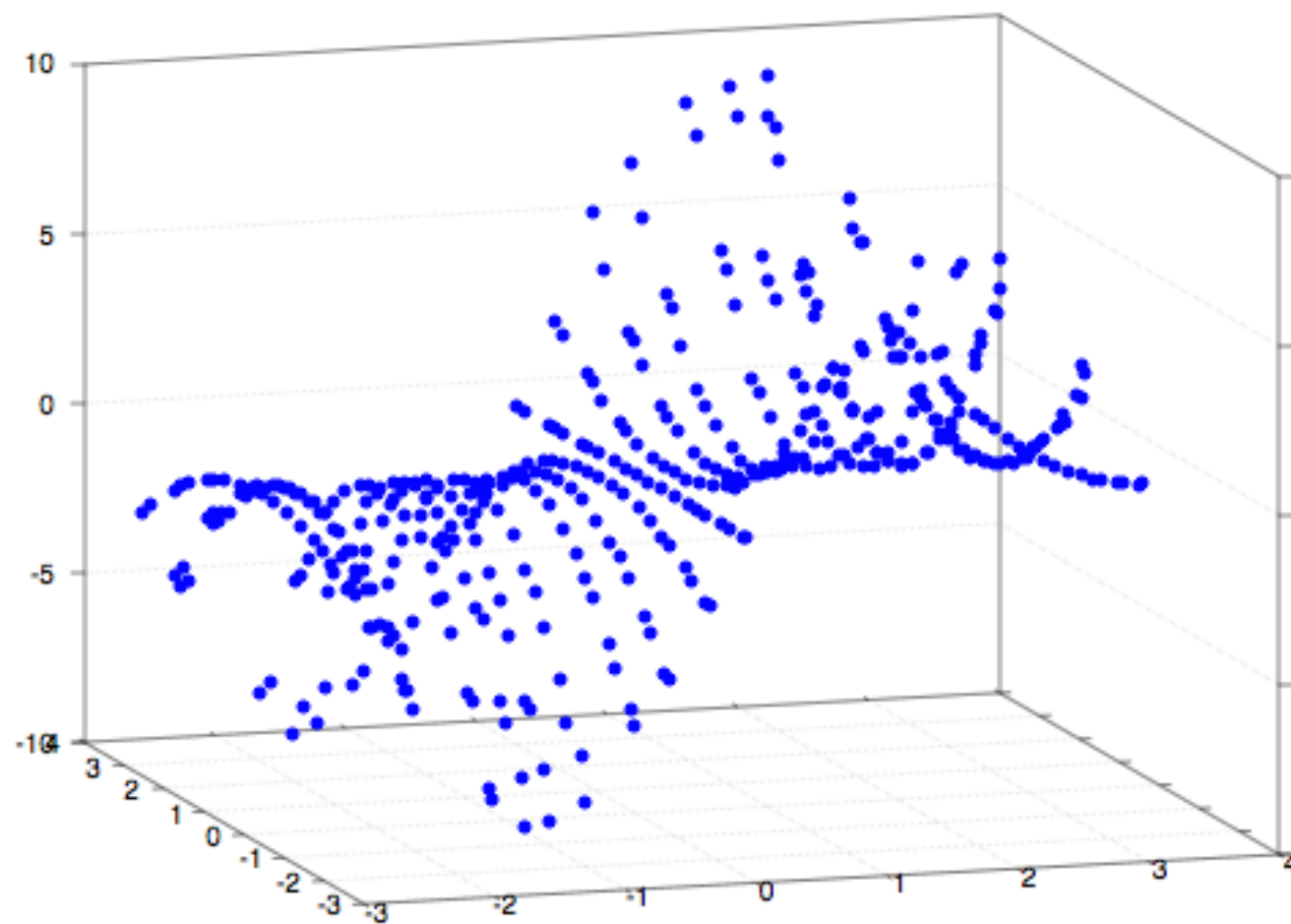
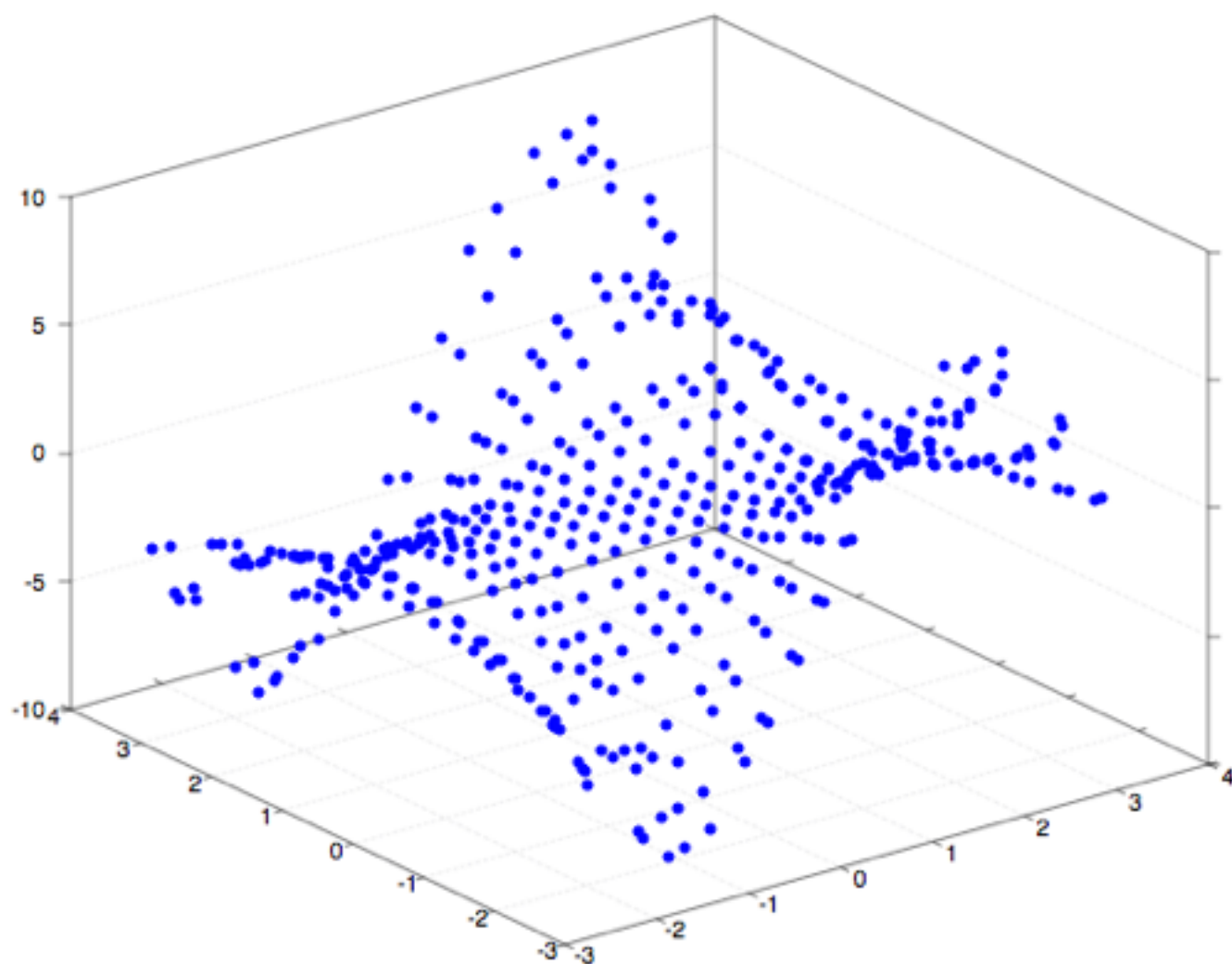
Julián Gutierrez
Alexis Medvedeff
Javier Perez Cuñarro

Agenda

1. Problema a resolver
2. Variables
3. Resultados
4. Conclusiones

Problema

Estimar una función escalar
a partir de un conjunto
de puntos.



Conjunto de puntos
samples8.txt

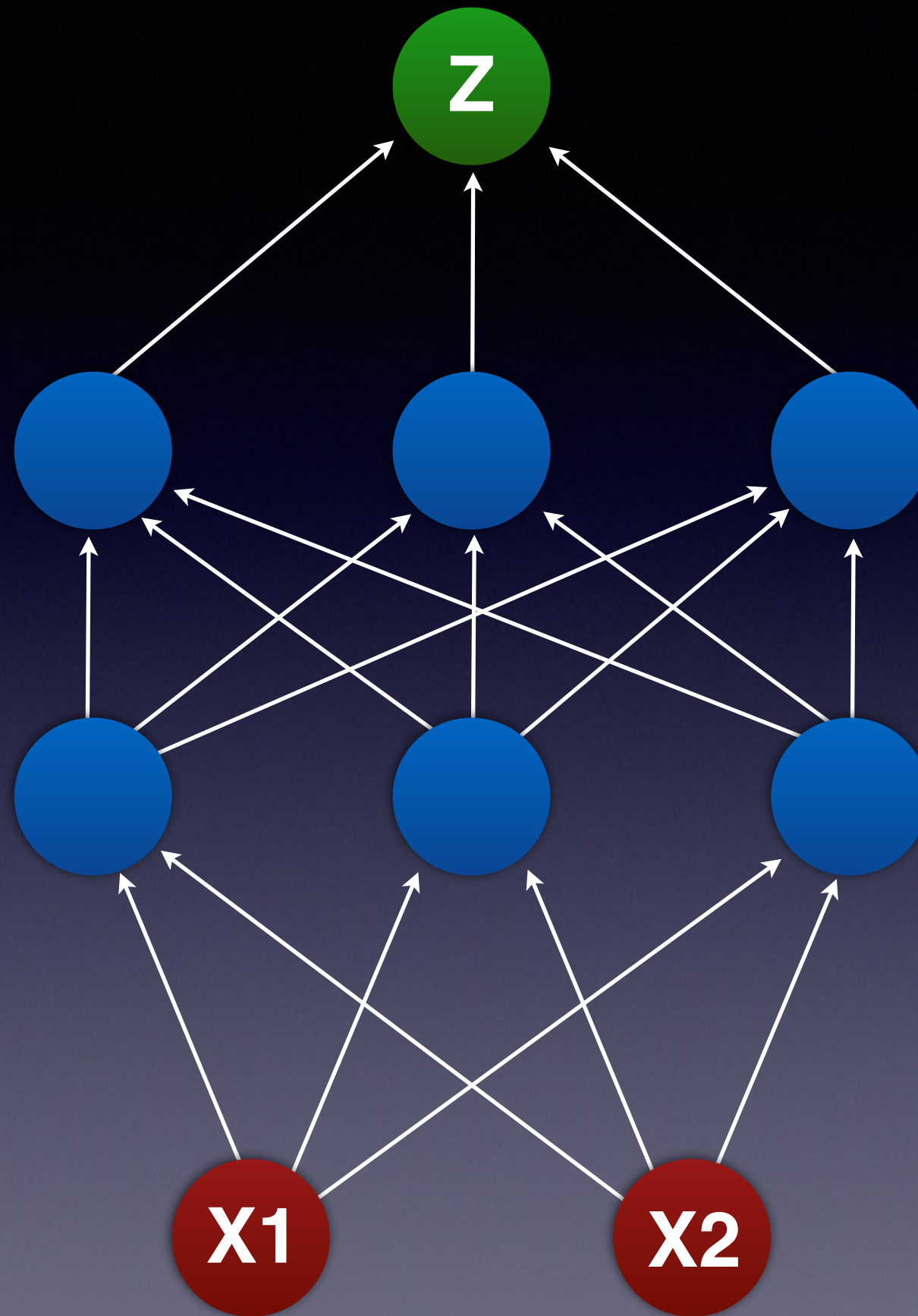
Solución

Implementar un
perceptrón multicapa con
aprendizaje supervisado.

Salida

Capas
ocultas

Entradas



Identificamos la red como **[3 3]**

¿Configuración óptima?

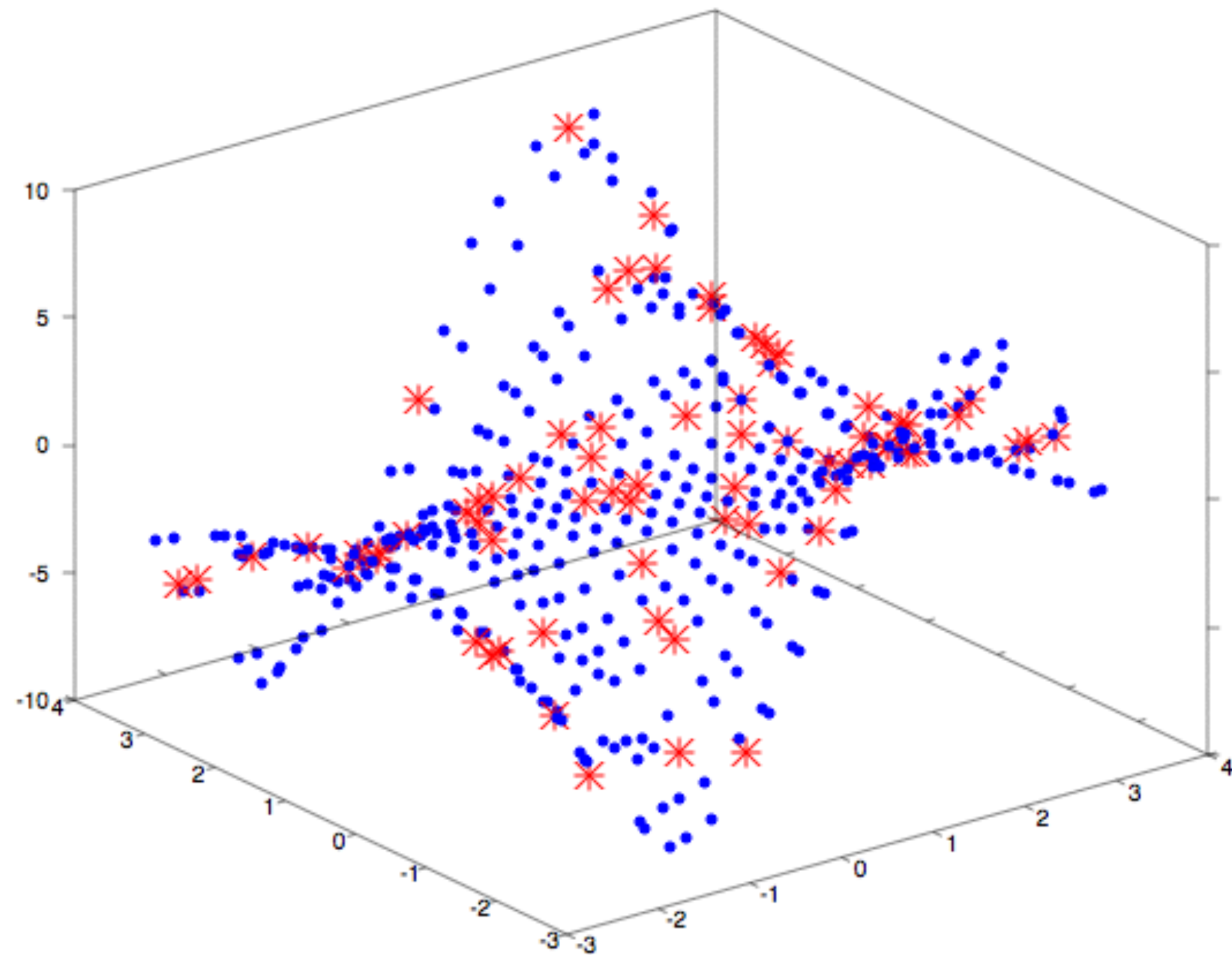
Resultados

Variable 1

Cómo **dividir los puntos de entrada** para las fases de entrenamiento y de testeo.

Comparación entre porcentajes

División de datos de entrada	Error cuadrático medio, promedio en testeo	Porcentaje de aciertos con error menor a 10^{-3}
40% entrenamiento 60% testeo	0.8478%	48.6013%
60% entrenamiento 40% testeo	0.9522%	44.9101%
80% entrenamiento 20% testeo	0.7181%	41.1764%



Fase de **entrenamiento**: 0.8

Fase de **testeo**: 0.2

Variable 2

Qué **función de activación**
es más apropiada.

Función **tangente** **hiperbólica**

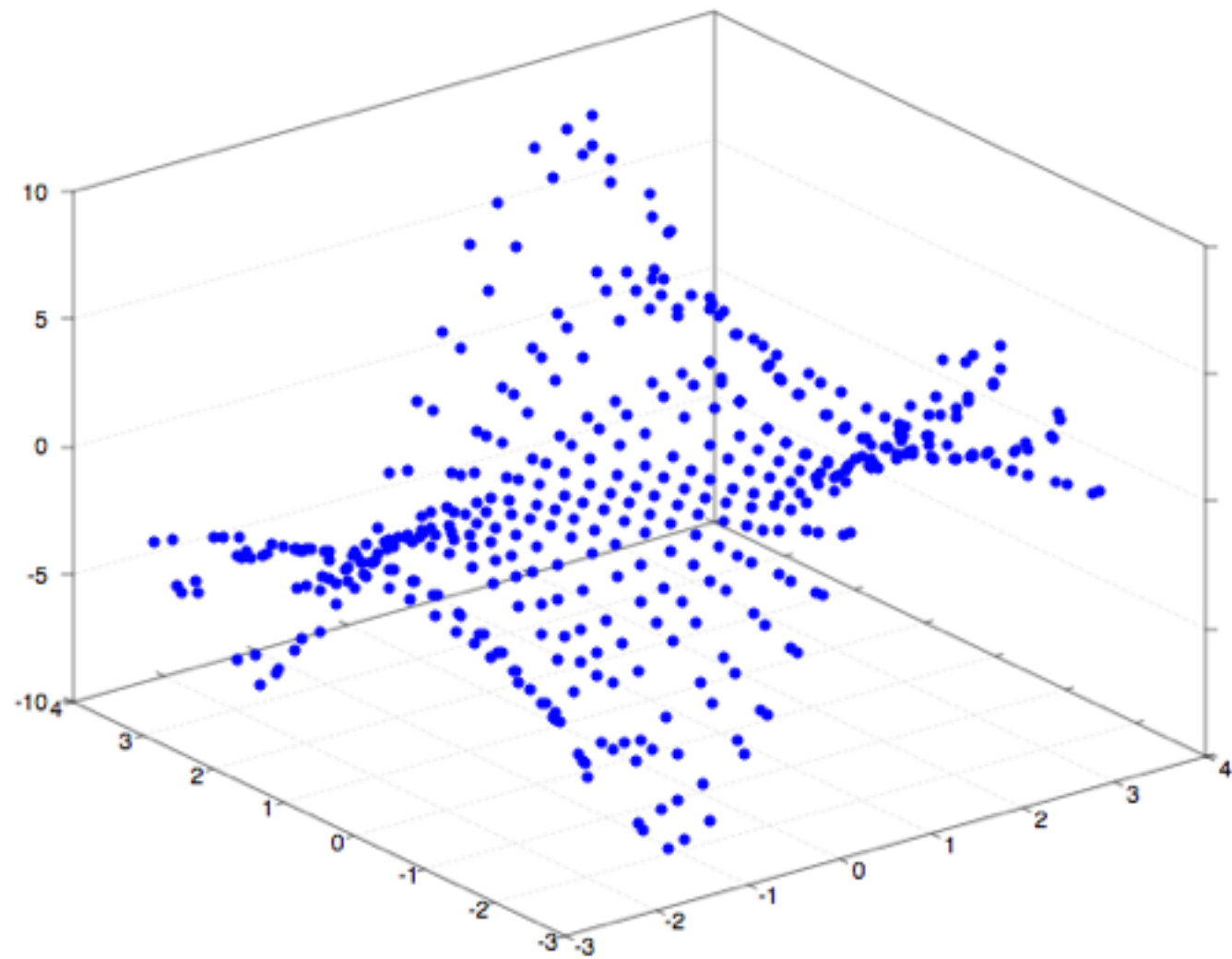
$$f(x) = \tanh(\beta x) \quad \beta = 1$$

Función **exponencial**

$$g(x) = \frac{1}{1 + e^{(-2\beta x)}} \quad \beta = 2$$

Comparación entre funciones

Función de activación	Error cuadrático medio, promedio en testeo	Porcentaje de aciertos con error menor a 10^{-3}
tanh beta=1	1.2850%	28.6863%
tanh beta=0.5	1.4006%	17.6470%
exp beta=1	1.3994%	32.6921%
exp beta=2	0.7794%	35.3346%



Comparación entre tangente
hiperbólica y exponencial

Variable 3

Con qué **arquitectura de red** aprende mejor el perceptrón.

Comparación entre arquitecturas

Arquitectura	Error cuadrático medio, promedio en testeo	Porcentaje de aciertos con error menor a 10^{-3}
[20]	1.2924%	36.7647%
[15 15]	0.9339%	30.8823%
[10 20 30]	0.7181%	41.1764%
[2 4 8 16]	1.5001%	25.0000%

Variable 4

Cuáles de las **mejoras al backpropagation** impactan positivamente en los resultados.

Mejora η adaptativo

$$\Delta\eta = \begin{cases} +a & \text{si } \Delta E < 0 \text{ consistentemente} \\ -b\eta & \text{si } \Delta E > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$a = 2,5\%$ del η inicial

$b = 25\%$ del η inicial

se modifica η cada 5 épocas

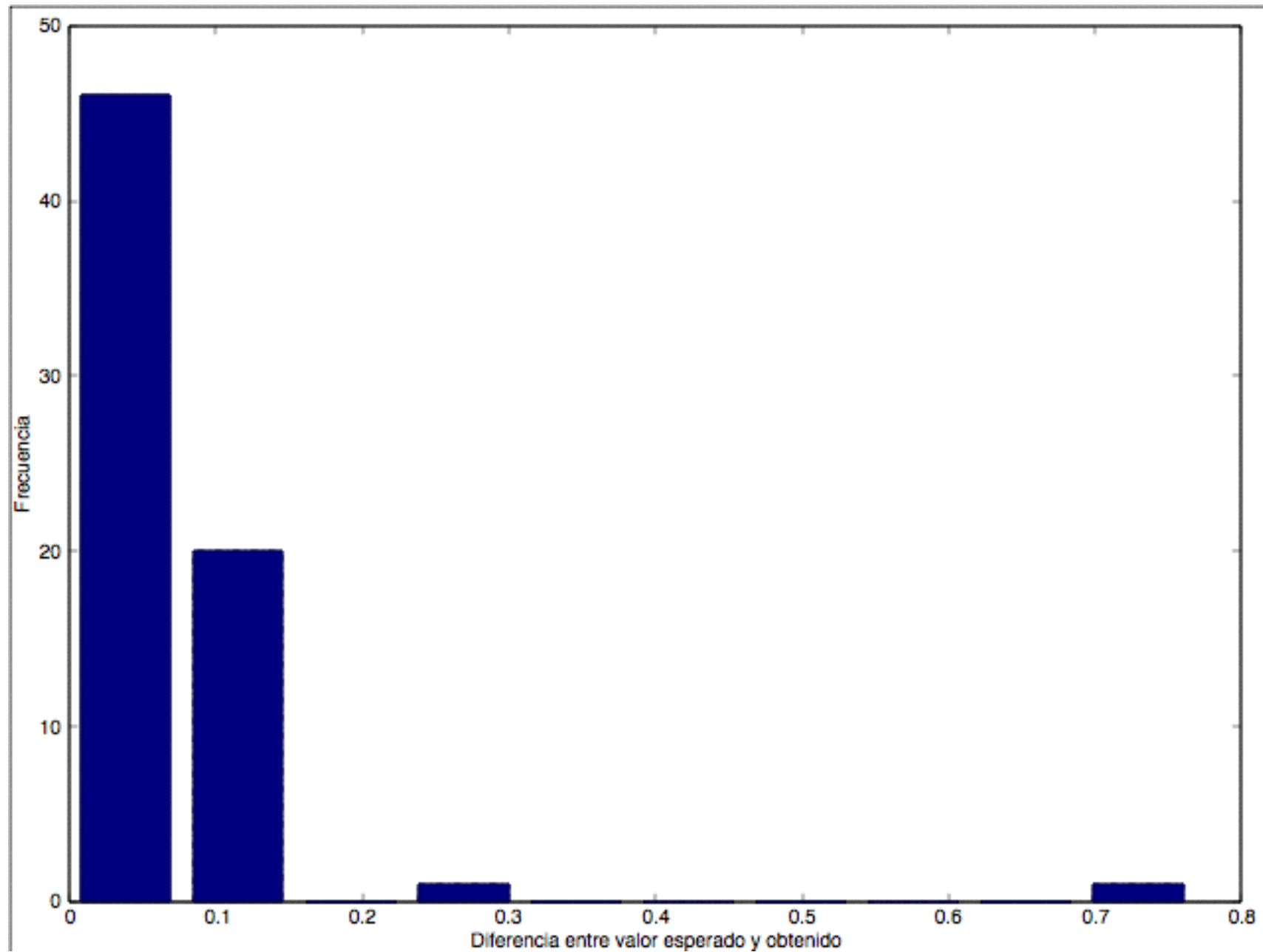
Mejora momentum

$$\Delta w_{pq}(t + 1) = -\eta \frac{\partial E}{\partial w_{pq}} + \alpha \Delta w_{pq}(t) \quad 0 < \alpha < 1$$

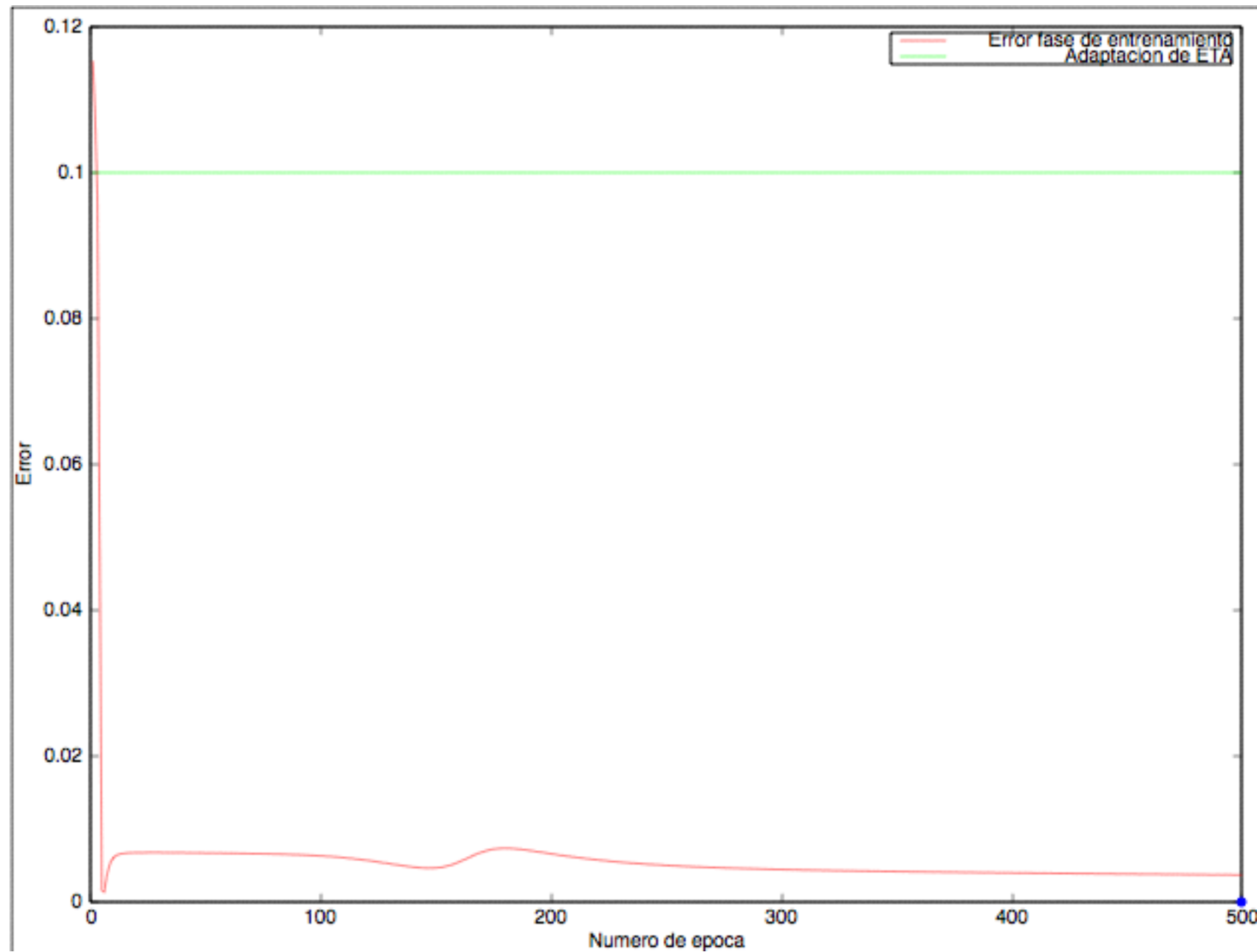
$$\alpha = 0.9$$

Comparación para backtracking

Mejoras aplicadas	Error cuadrático medio, promedio en testeo	Porcentaje de aciertos con error menor a 10^{-3}
Ninguna	1.5060%	26.4705%
Momentum	1.6598% ($\alpha=0.9$)	26.4705% ($\alpha=0.9$)
	1.5405% ($\alpha=0.1$)	26.4705% ($\alpha=0.1$)
η adaptativo	1.5187%	17.6470%
Momentum + η adaptativo	2.2648% ($\alpha=0.9$)	8.8235% ($\alpha=0.9$)
	1.6079% ($\alpha=0.1$)	19.1176% ($\alpha=0.1$)



Diferencia valor calculado y esperado
Red [10 20 30]



Error cuadrático por época
Red [10 20 30]

Conclusiones

La decisión de la cantidad de neuronas y capas no es decidible de forma teórica.

Es necesario realizar pruebas para ver si realmente conviene una forma o no por sobre otra. Lo mismo sucede con los distintos parámetros como puede ser el momentum o el eta y los parámetros de adaptación del mismo.

Conclusiones

La selección de la arquitectura y los parámetros depende de las condiciones del problema.

Conclusiones

Utilizar la función de activación exponencial puede llevar en algunos casos a estancarse en mínimos.

