



Acceleration of a bacterial metabolic simulation using neural networks with optimization techniques

TRABAJO DE FIN DE MÁSTER

MÁSTER EN INTELIGENCIA ARTIFICIAL – UNIVERSIDAD POLITÉCNICA DE MADRID

AUTOR:

Javier París Uhryn

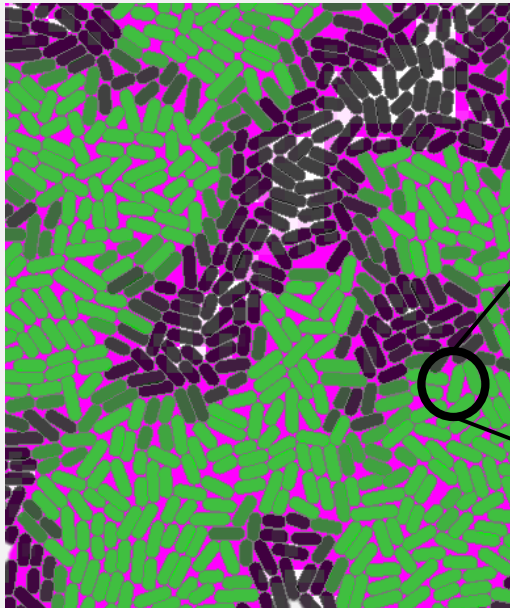
TUTOR:

Alfonso Rodríguez-Patón Aradas

Acceleration of a bacterial metabolic simulation using neural networks with optimization techniques

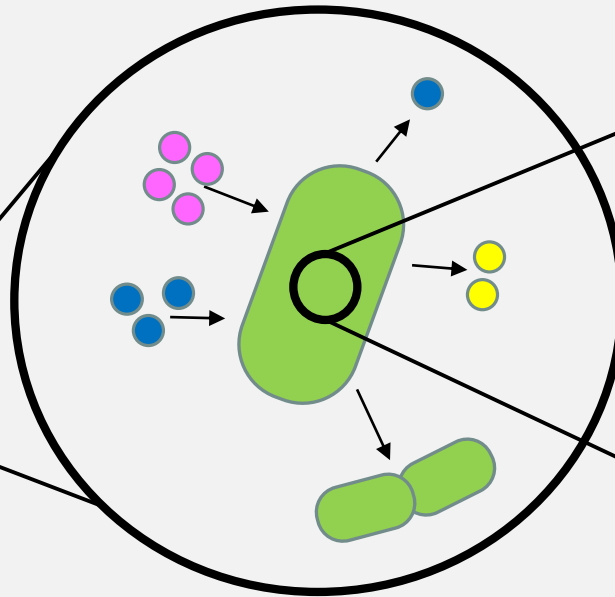
GRO

Simulador basado en
agentes para simular
ecosistemas
bacterianos



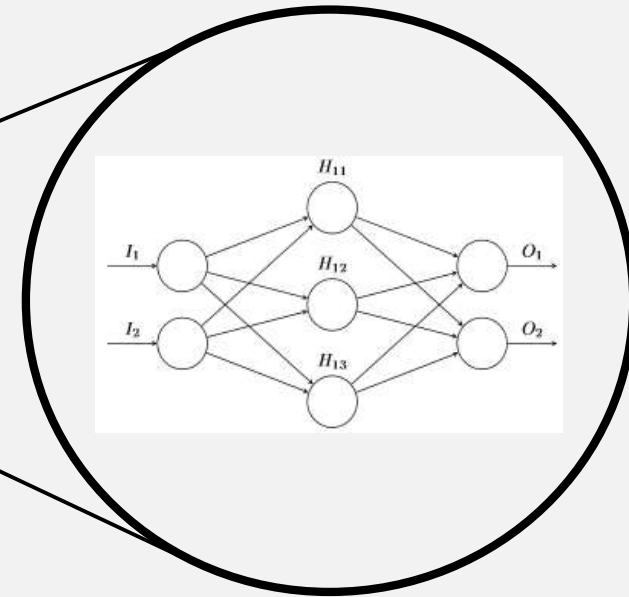
Metabolismo

Individual, flexible
y realista

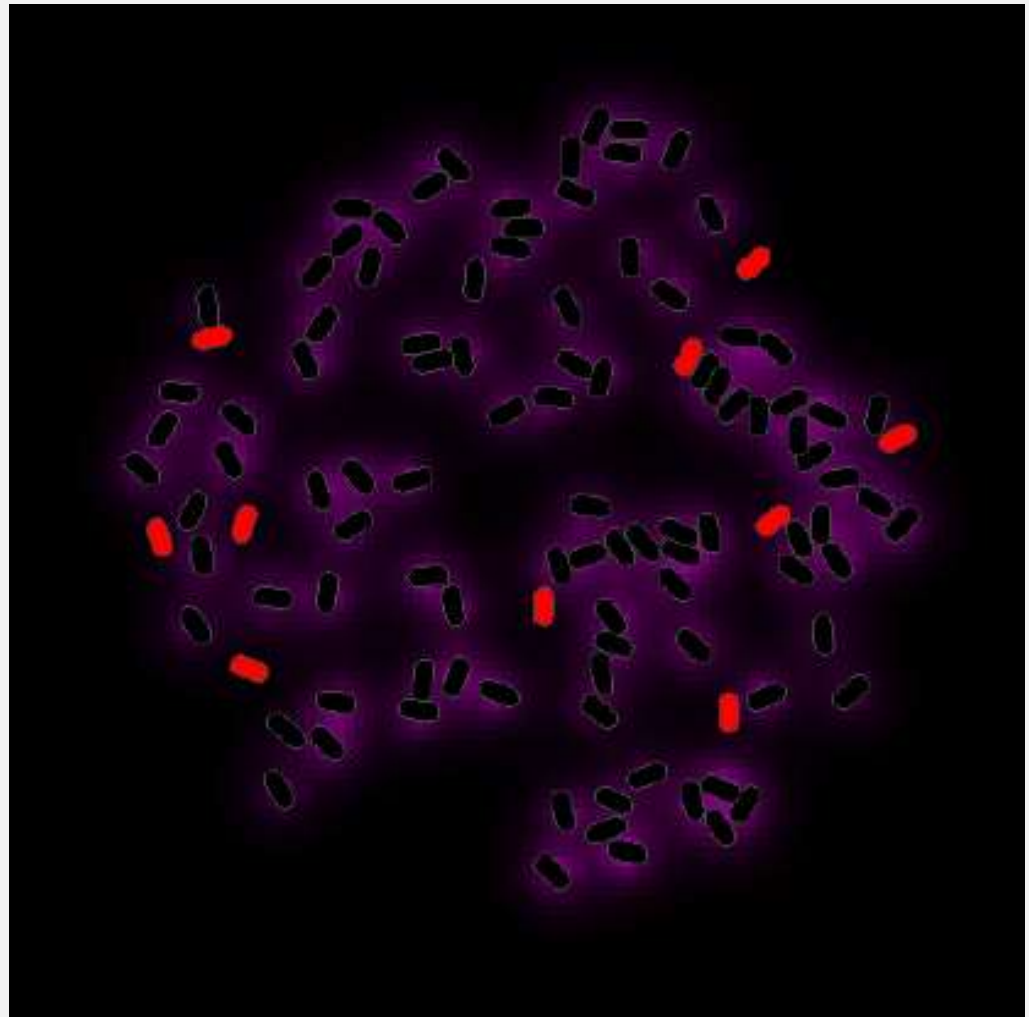


Redes neuronales

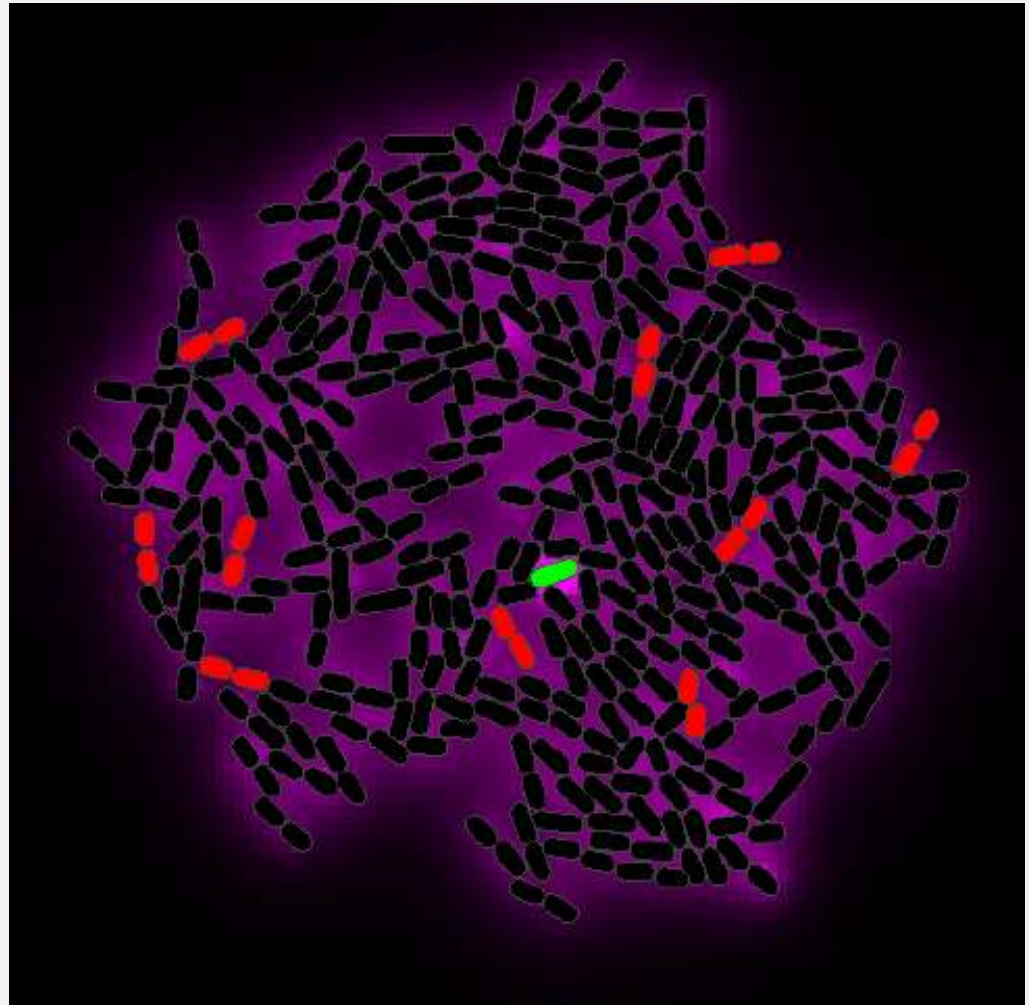
Aproximación
rápida del
metabolismo



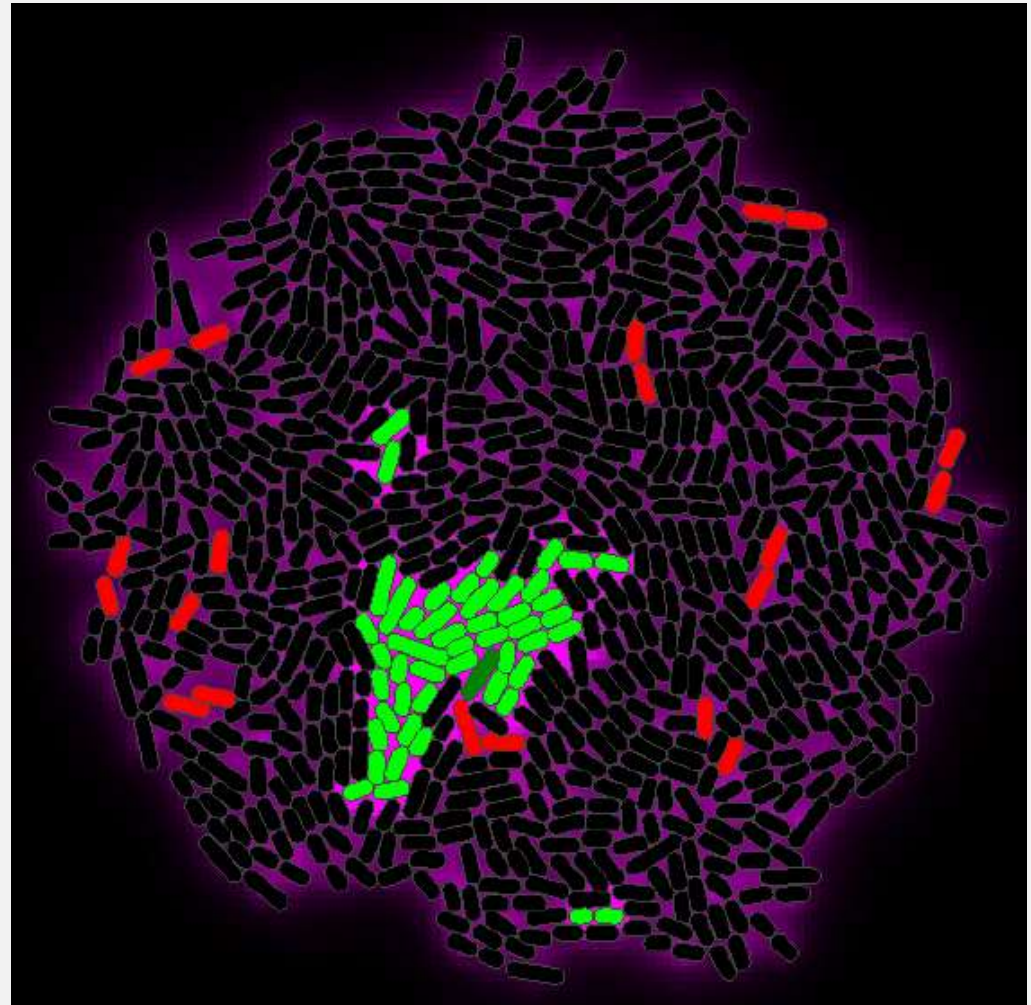
- Simulador basado en individuos/agentes



- Simulador basado en individuos/agentes
- Las interacciones entre individuos mediante señales

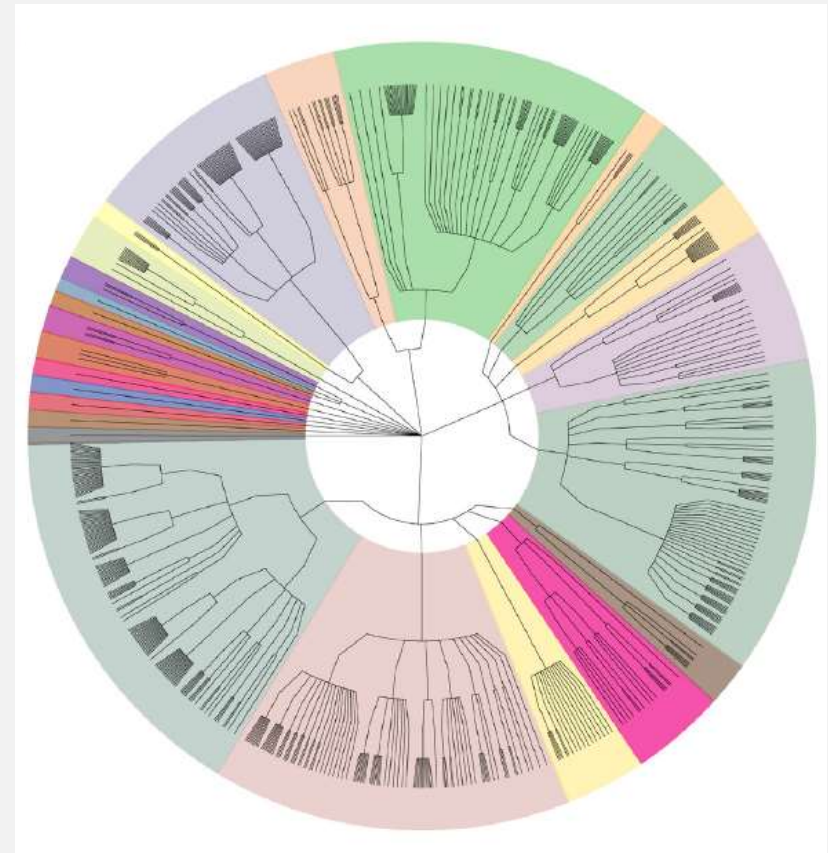
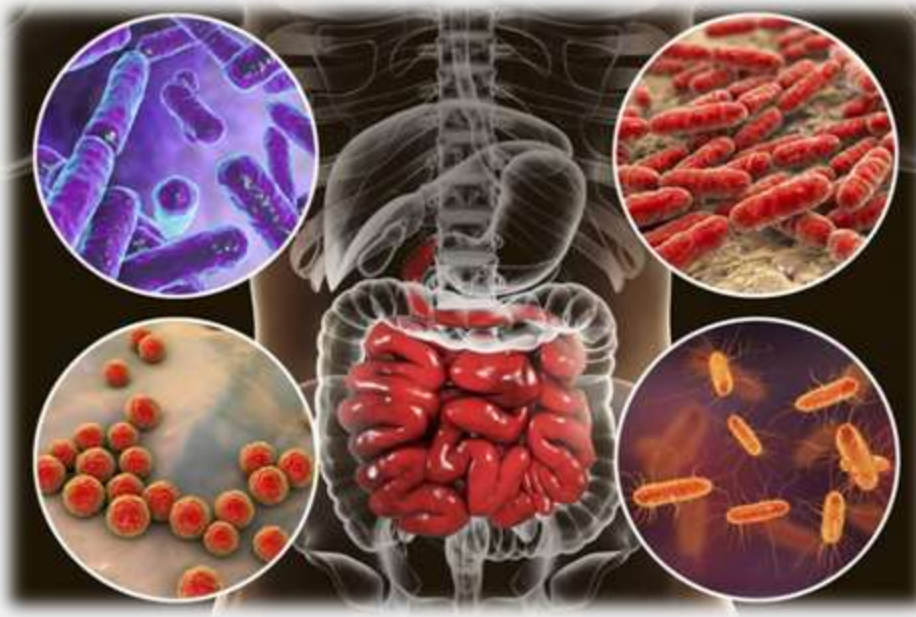


- Simulador basado en individuos/agentes
- Las interacciones entre individuos mediante señales
- La disposición espacial afecta al comportamiento de la colonia



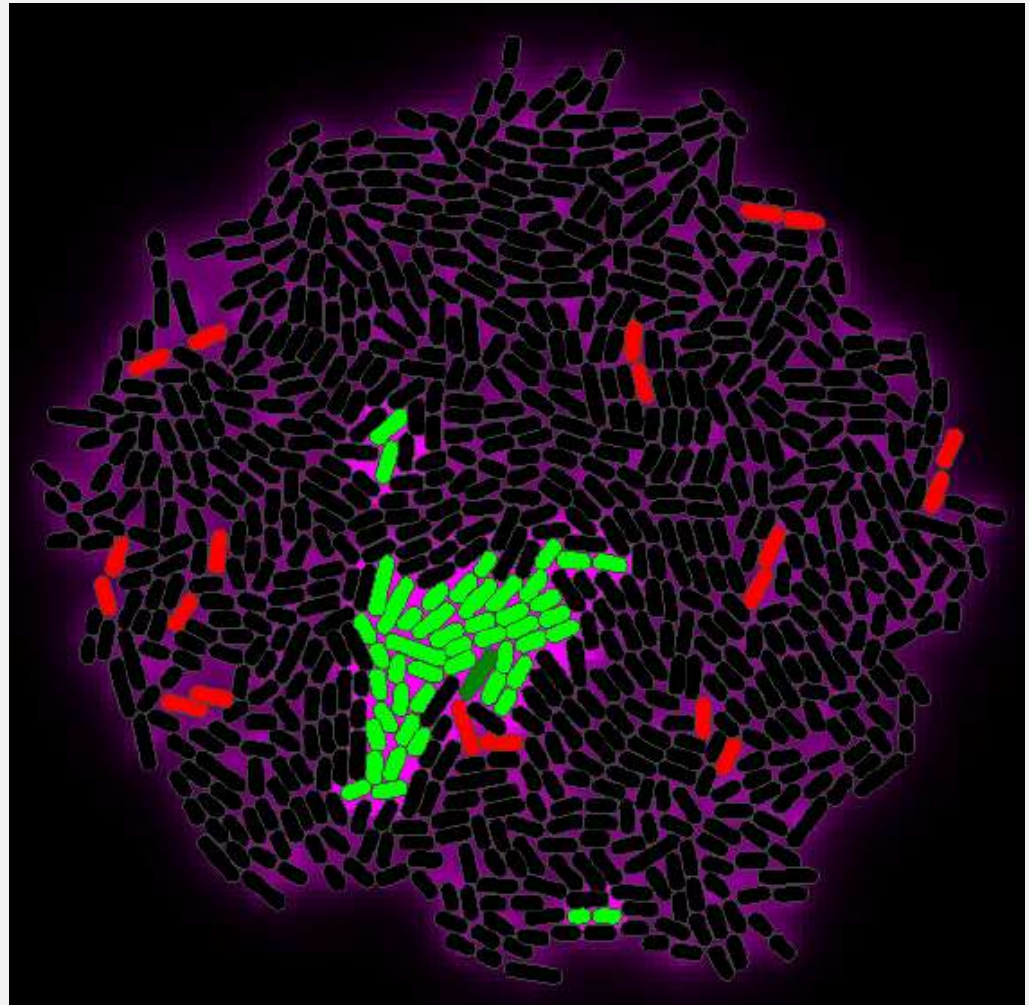
Es interesante estudiar consorcios bacterianos complejos

25 clases, 205 géneros y **605** especies bacterianas diferentes en la microbiota humana



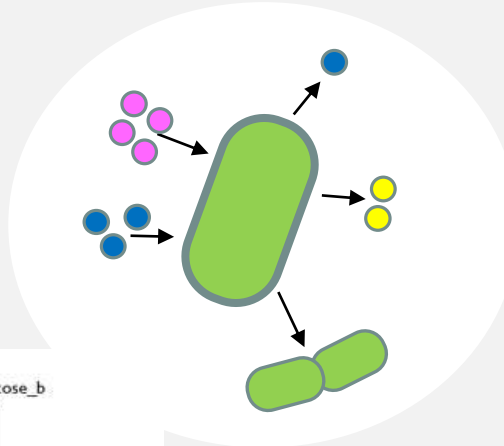
Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota, natura biotechnology

- Simulador basado en individuos/agentes
- Las interacciones entre individuos mediante señales
- La disposición espacial afecta al comportamiento de la colonia
- Se necesita mejorar el módulo de nutrientes
 - Flexible
 - Realista
 - Rápido

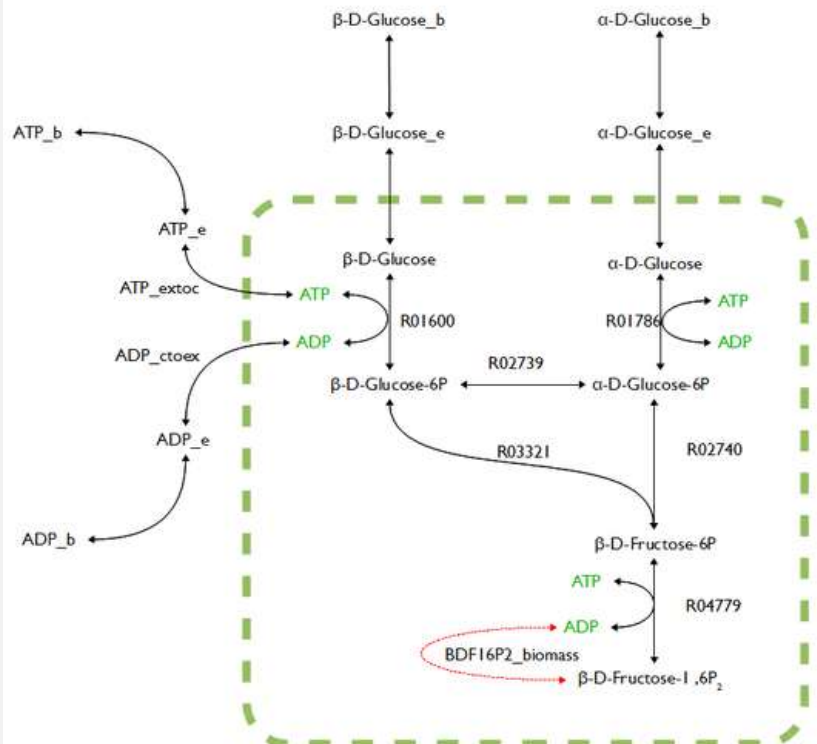


Metabolitos absorbidos y secretados y crecimiento asociado

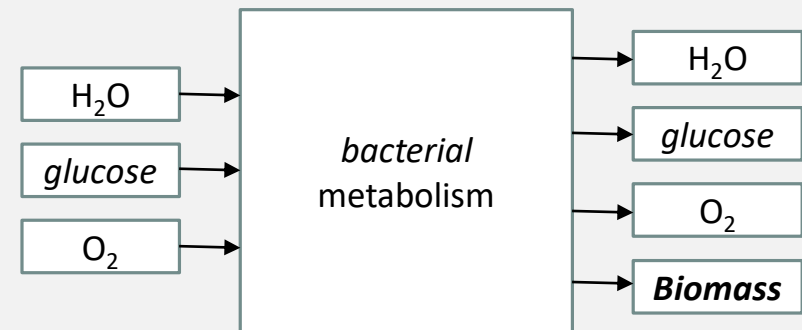
Solo nos interesa las entradas y salidas



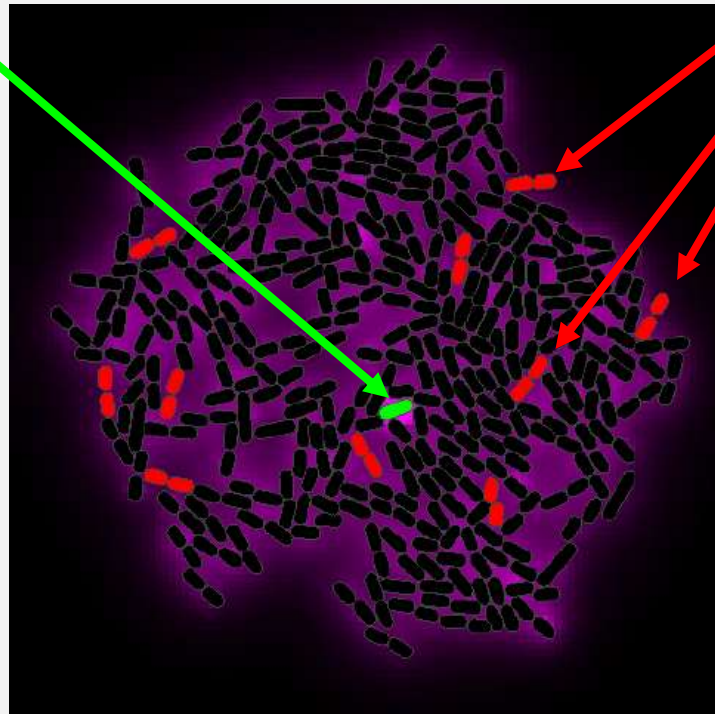
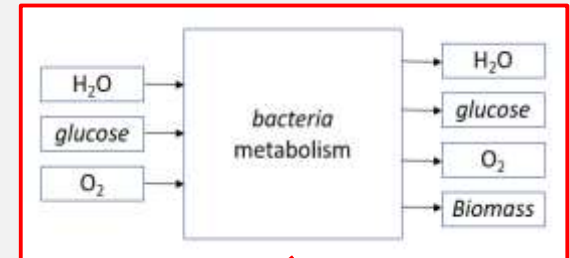
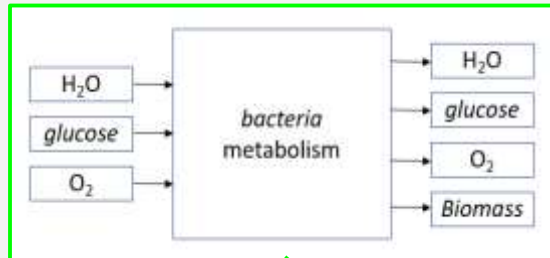
Modelo metabólico

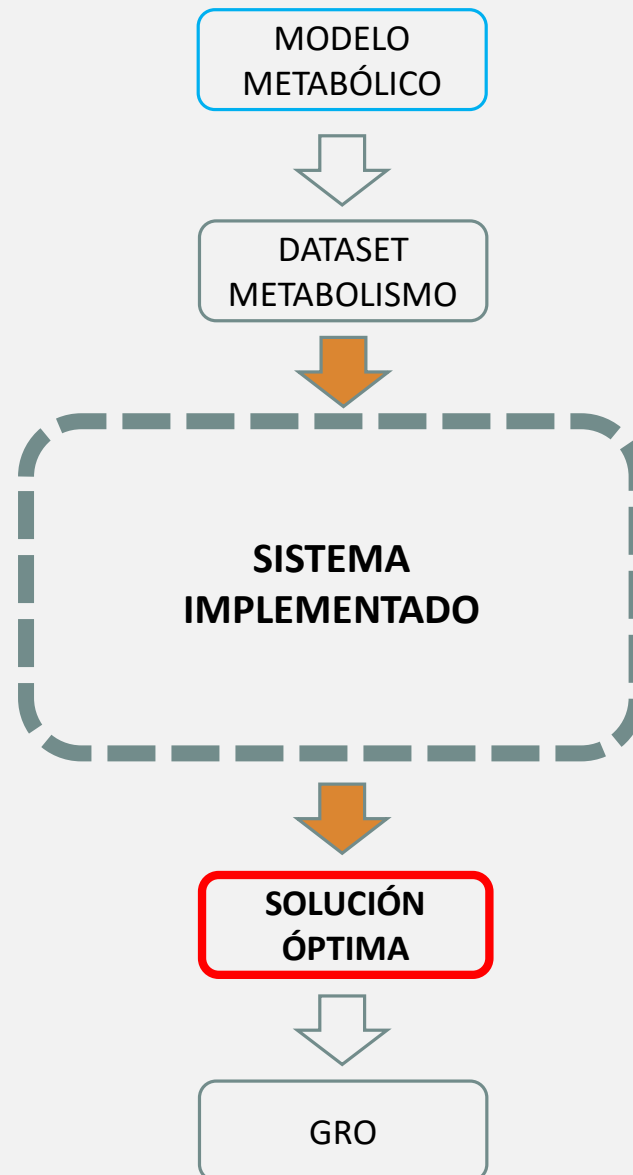


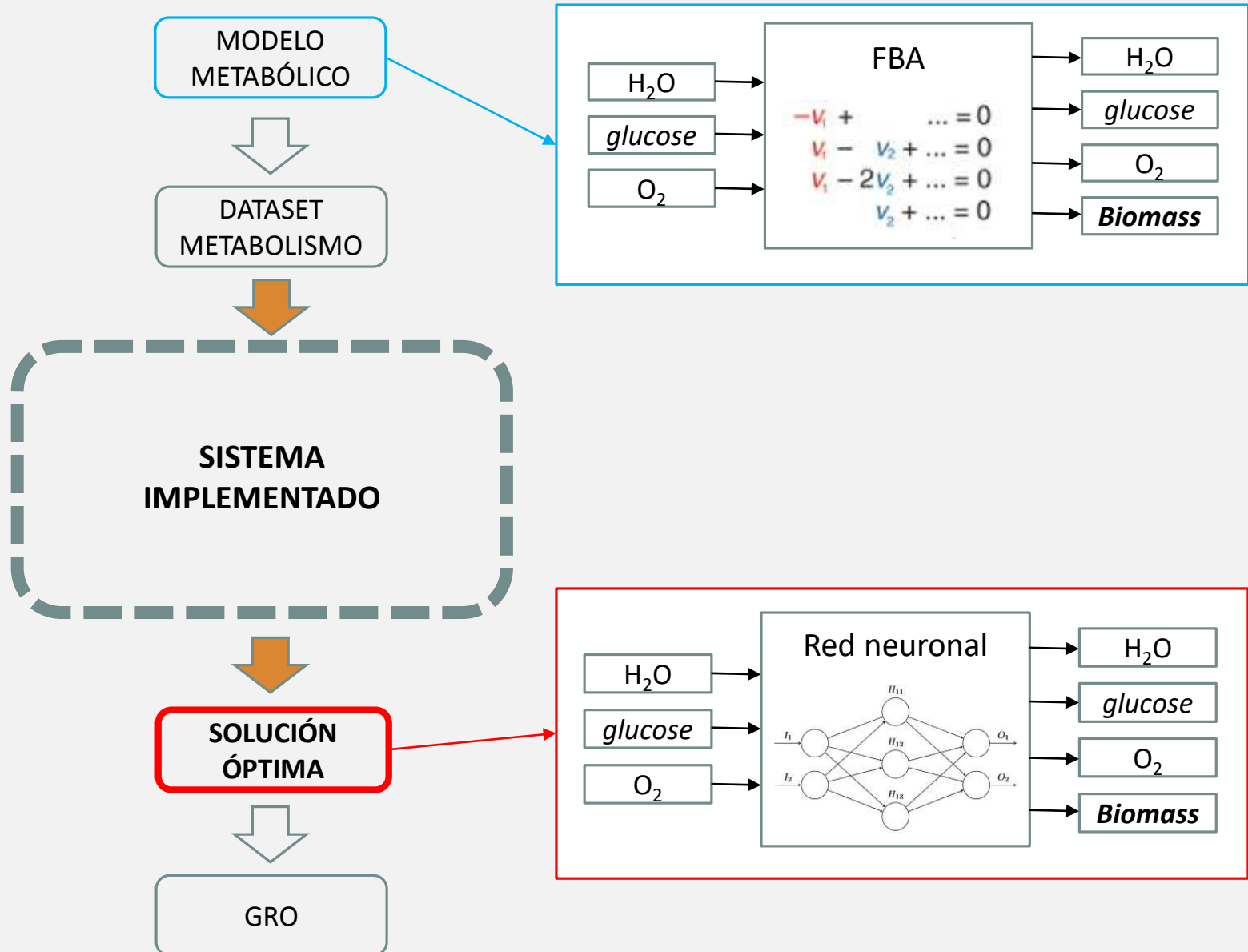
Caja negra como función metabólica



Cada célula tendrá un metabolismo individual dependiendo de su especie y cepa







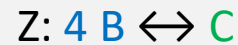
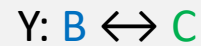
IDEA PRINCIPAL	CONCEPTOS	IMPLEMENTACIÓN	CONCLUSIONES
GRO	Metabolismo		Solución

HIPÓTESIS

- **H1:** FBA es aproximable con precisión mediante una red neural
 - 10% de error máximo
- **H2:** La red neuronal que aproxime FBA deberá ser ejecutada rápidamente
 - Mínimo de 100 veces más rápido que el FBA
- **H3:** La red que satisfaga **H1** y **H2** para un *dataset* dado debe ser encontrada en un tiempo límite
 - Máximo tiempo de respuesta de 10 minutos

Flux Balance Analysis (FBA)

Reacciones



Matrix estequiométrica

Metabolites	A (ext)	B (ext)	X	Y	Z	Biomass
A	-1		-1			
B		-1	-1	-1	-4	
C			2	1	1	1

Sistema de ecuaciones

$$-A(\text{ext}) - X = 0$$

$$-B(\text{ext}) - X - Y - 4Z = 0$$

$$X + Y + Z + \text{Biomass} = 0$$

Restricciones

$$Y < 10$$

$$\text{Biomass} < 25$$

Flux Balance Analysis (FBA)

Reacciones

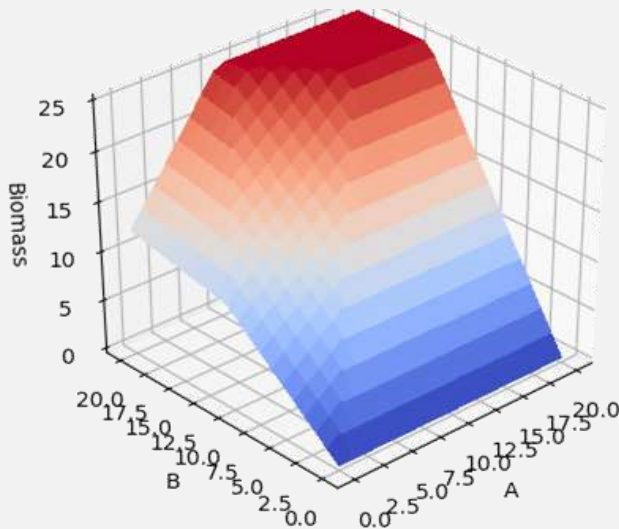


Restricciones

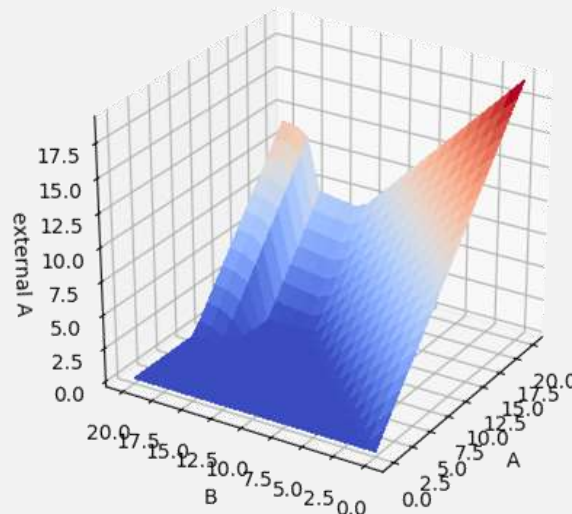
$$Y < 10$$

$$\text{Biomass} < 25$$

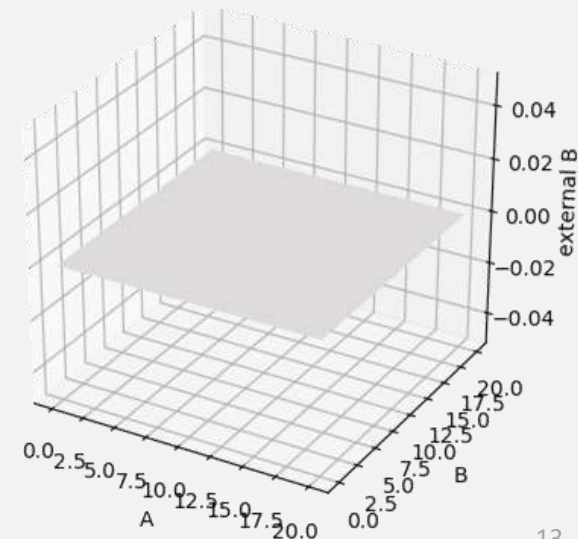
Biomass producida



A sobrante



B sobrante

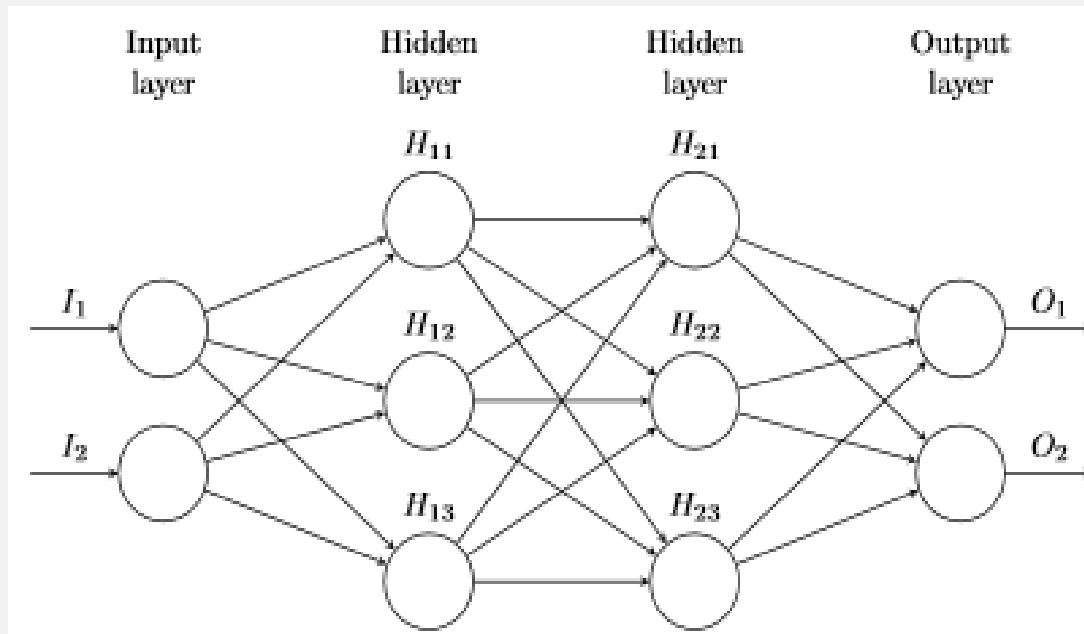


IDEA PRINCIPAL	CONCEPTOS	IMPLEMENTACIÓN	CONCLUSIONES
FBA	Redes de Neuronas Artificiales	Algoritmos de optimización	

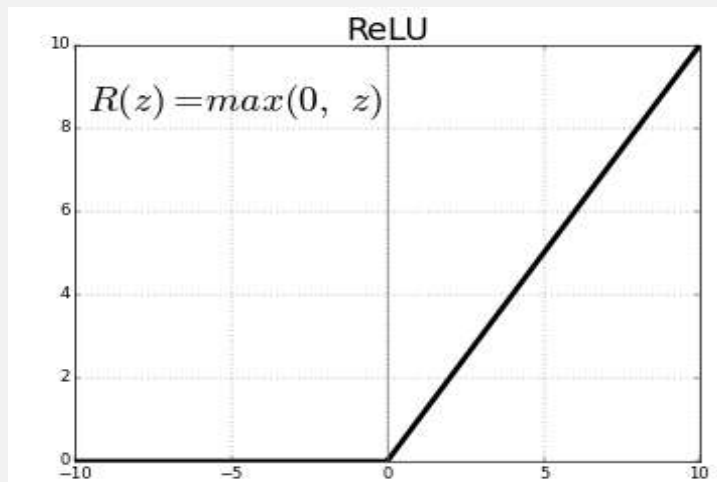
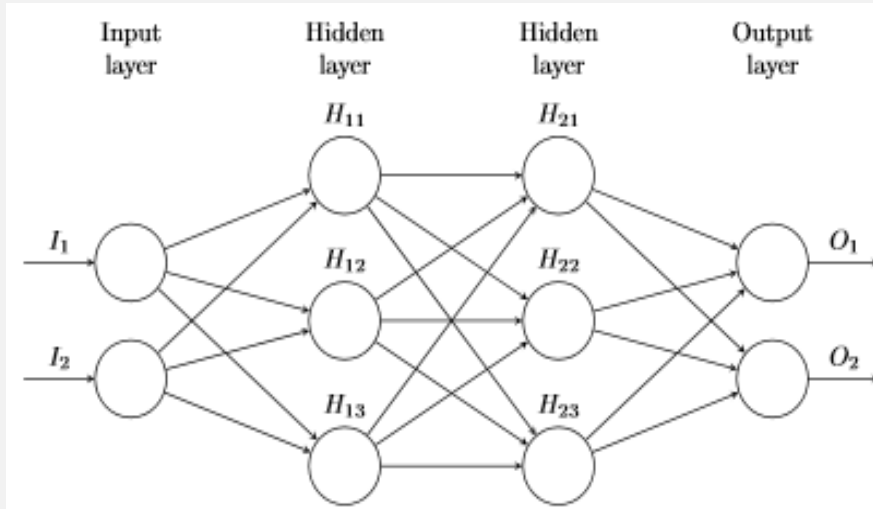
Modelos metabólicos utilizados

Modelo metabólico	Número de metabolitos	Número de metabolitos externos	Número de reacciones
<i>E. coli Core</i>	72	7	95
<i>E. coli</i>	1805	25	2583
<i>Salmonella</i>	1802	22	2546

Redes de Neuronas Artificiales

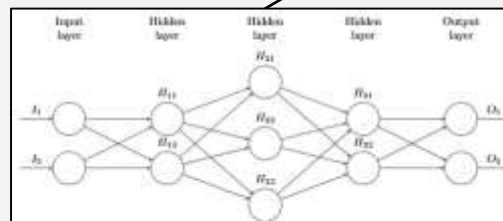
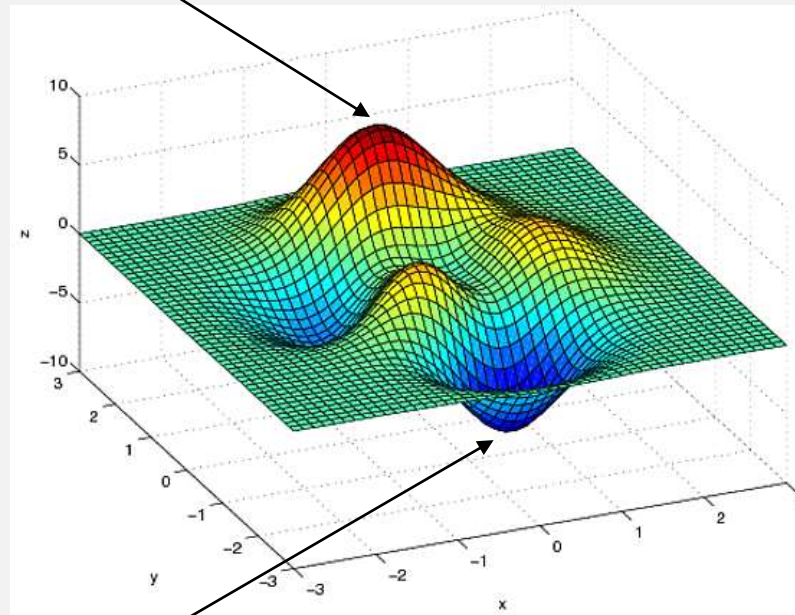
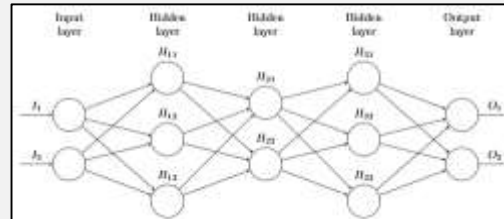


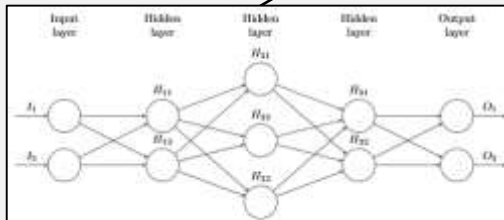
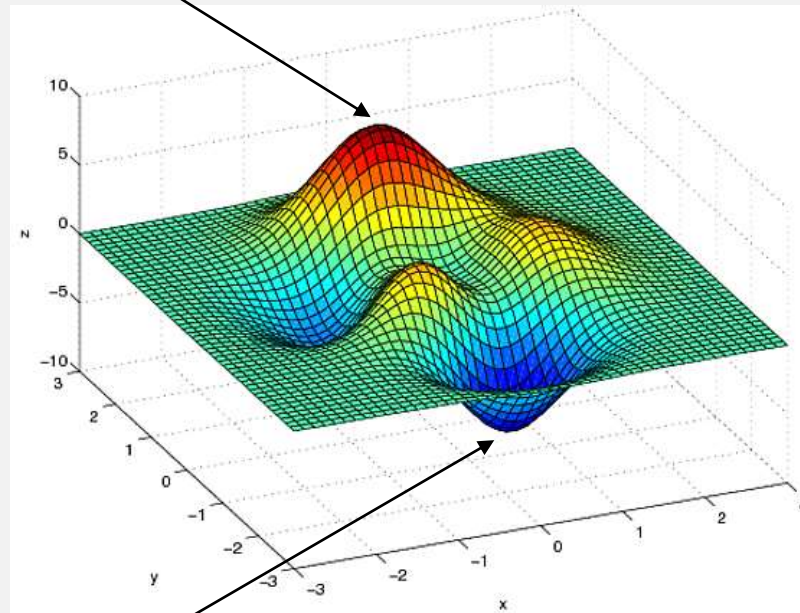
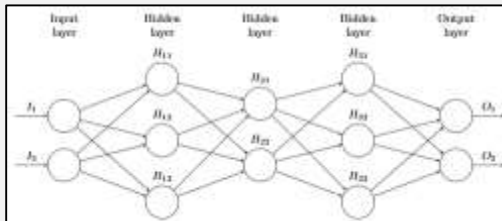
Diseño de la Red Neuronal A Priori



- **Red única**
 - Reduce el tiempo de ejecución
 - Las salidas están relacionadas
- **Feedforward fully connected**
 - Reduce la complejidad
 - Reduce la cantidad de estructuras diferentes
- **ReLU**
 - Es más rápida
 - Evita el problema del *desvanecimiento de gradiente*
 - Reduce el número de hiperparámetros
- **Resto de hiperparámetros**
 - Fijados para reducir complejidad
 - Solo varía la estructura

Optimización de la estructura de una red neuronal





Metaheurísticas:

- Colonia de hormigas
 - General
 - Por capas
- Recocido simulado
 - General
 - Creciente
- Algoritmos genéticos
 - Uniobjetivo
 - Multiobjetivo

Búsquedas exhaustivas:

- Creciente
- Decreciente
- Con diseño preestablecido
 - Diseño 2 capas
 - Diseño 3 capas

Diseño de experimentos

Hipótesis FBA

- **H1:** FBA es aproximable con precisión mediante una red neural
- **H2:** La red neuronal que aproxime FBA deberá ser ejecutada rápidamente
- **H3:** La red que satisfaga **H1** y **H2** para un *dataset* dado debe ser encontrada en un tiempo límite

Demostrar **H1** y **H2**

Se usa el dataset *E. coli*

Hipótesis optimización

Demostrar **H3**

Se usa el dataset *E. coli Core*

Solución final

Solución final alcanzada
cumpliendo **H1**, **H2** y **H3**

Se usan los 3 datasets:

E. coli Core

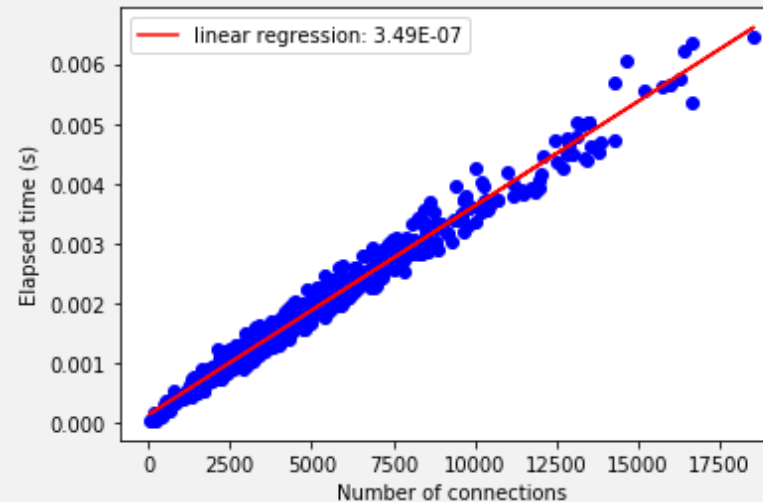
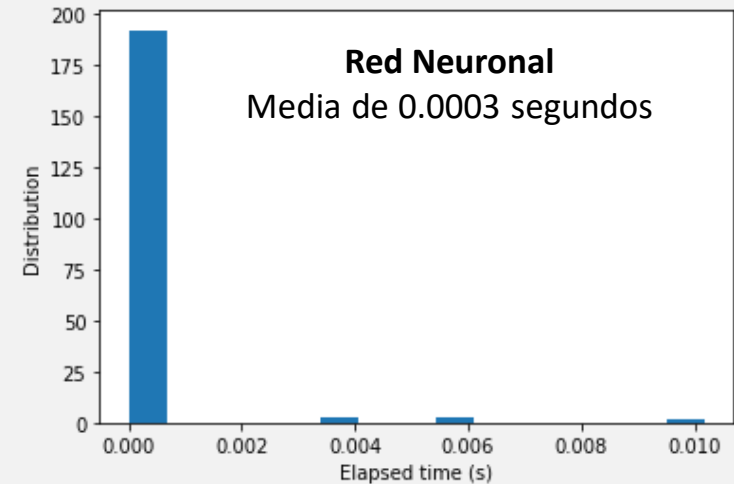
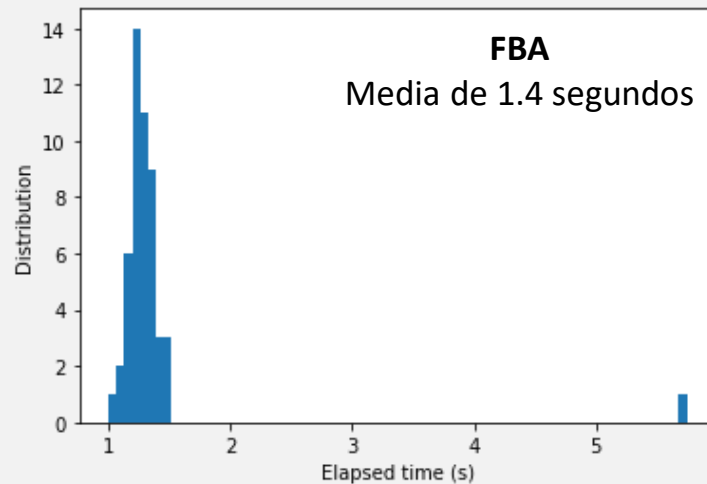
E. coli

Salmonella

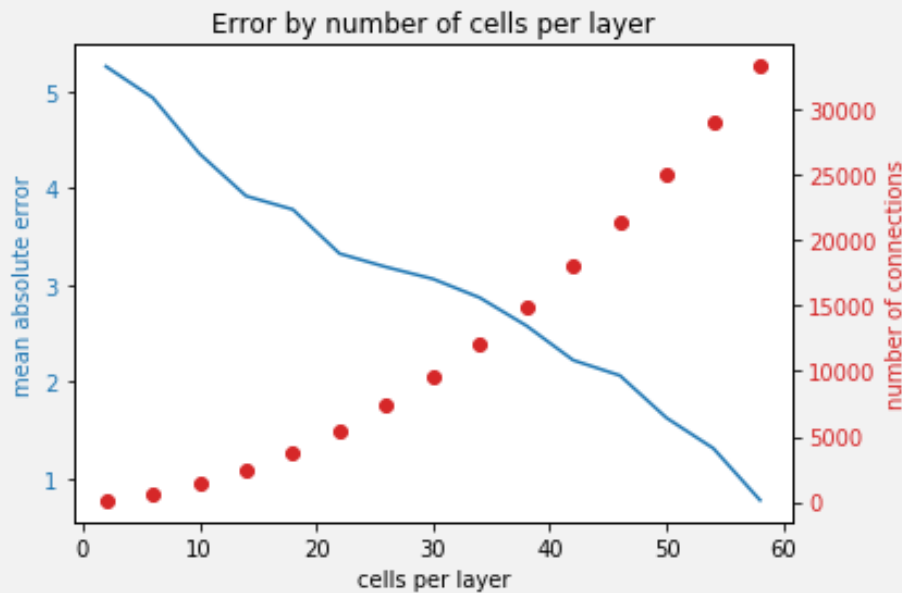
Tiempo de ejecución de FBA – *E. coli* frente a una red neuronal

El tiempo de ejecución de una red no varía con su input

Sí varía con su tamaño

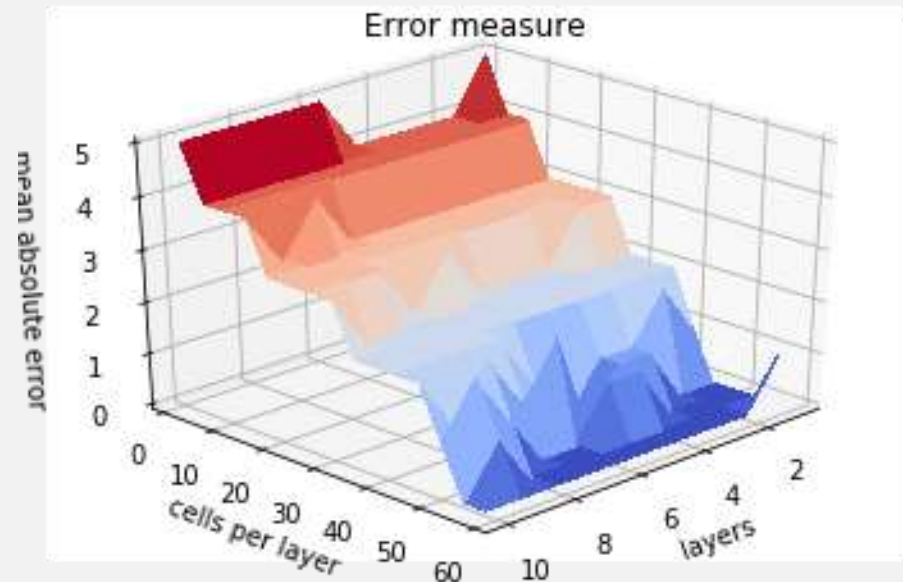


Error de una red aproximando FBA



El número de neuronas por capa es muy relevante en el error

Los mejores resultados se encuentran en redes poco profundas con muchas neuronas

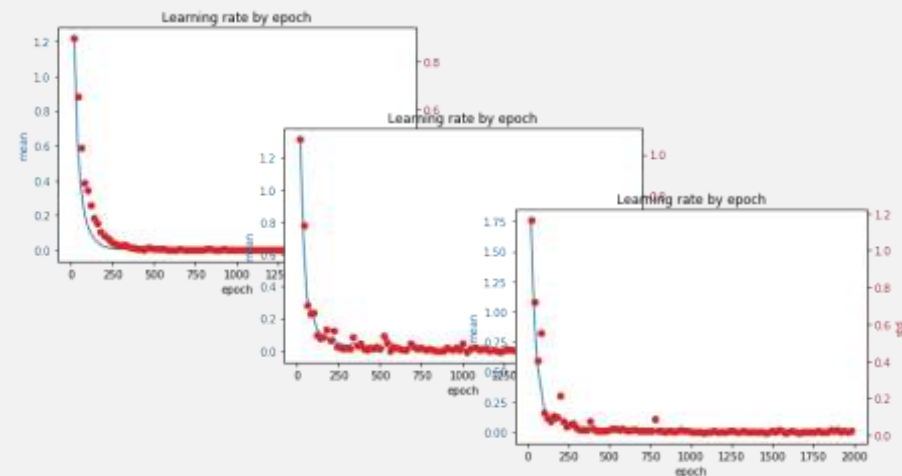
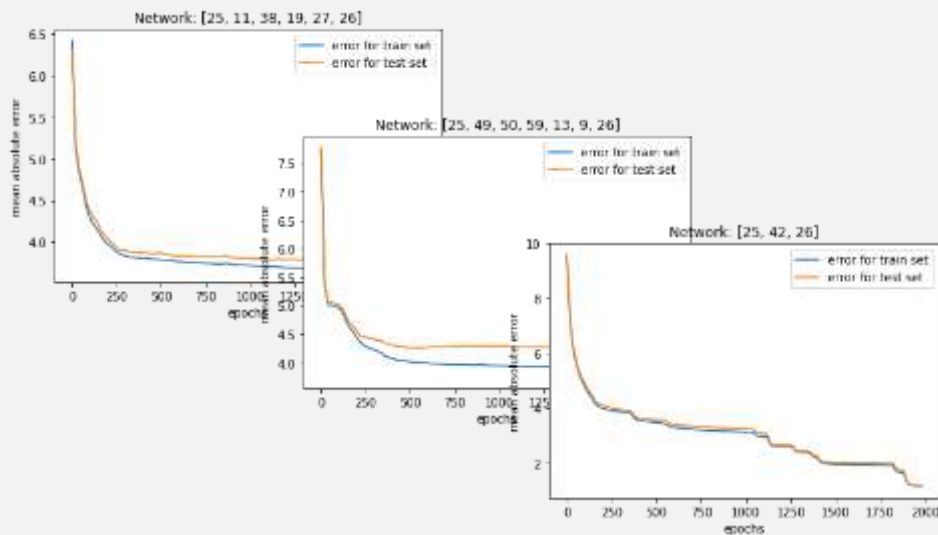
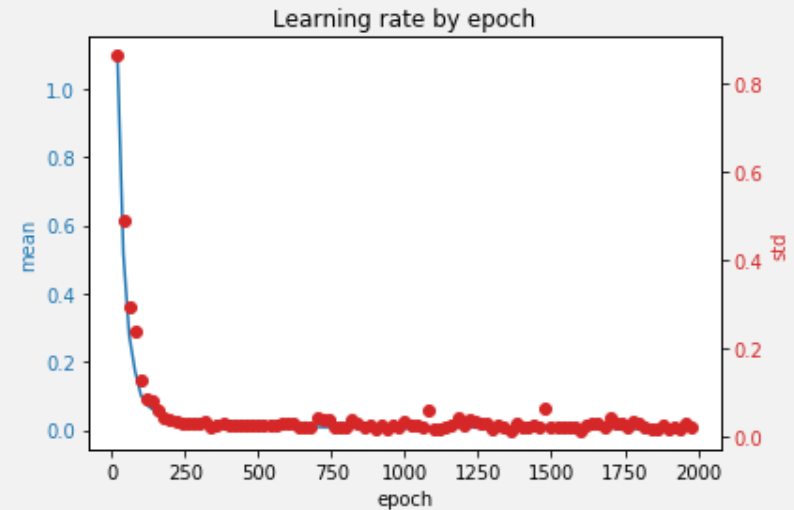


Evolución del error en entrenamiento

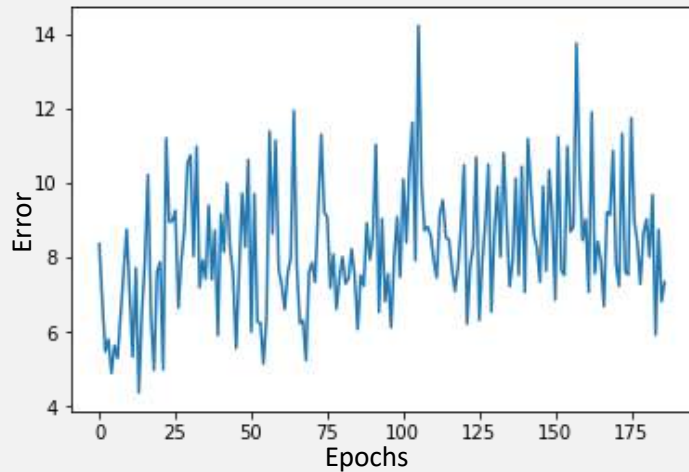
La curva de entrenamiento sigue una distribución similar en todas las redes



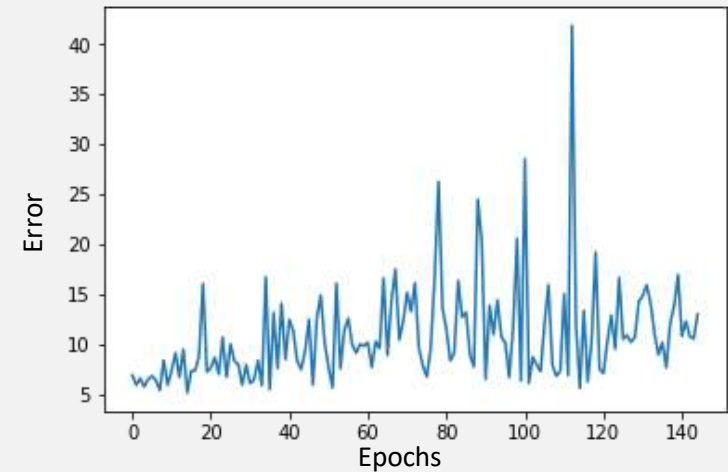
La evolución del error se reduce a las primeras épocas de entrenamiento



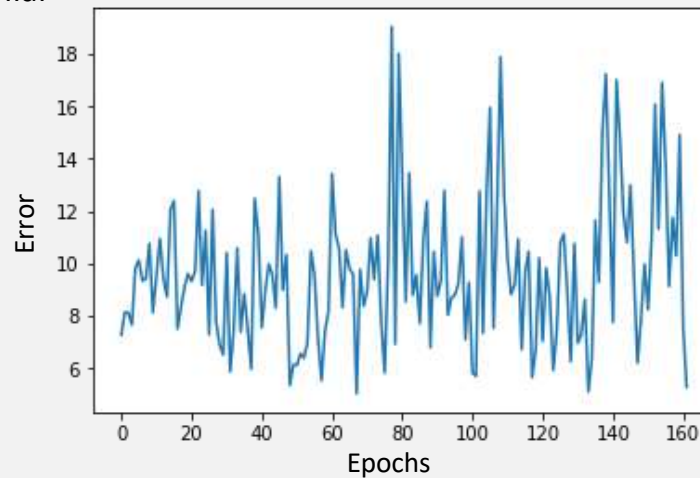
Resultados del Algoritmo de Colonia de Hormigas



Genérico - binomial

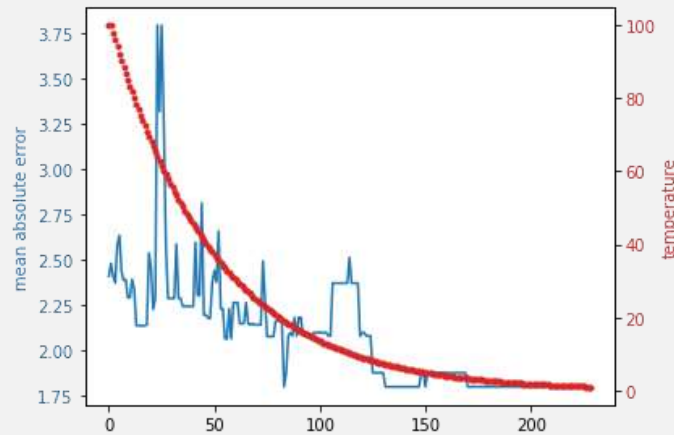


Genérico - geométrico

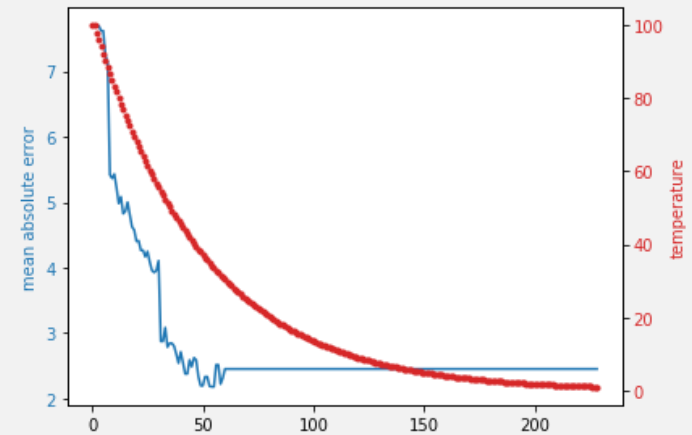


Por capa - gaussiano

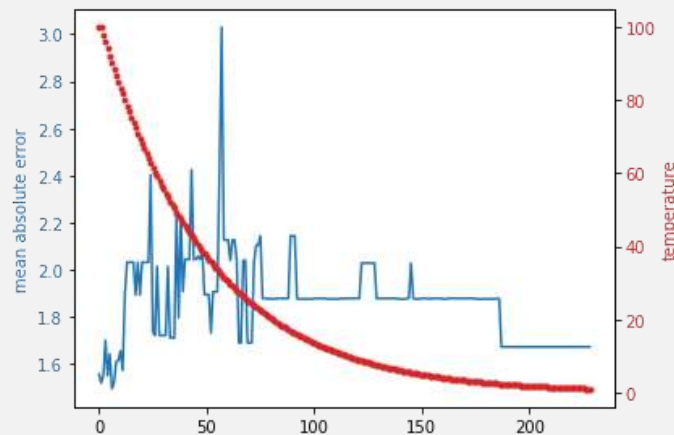
Resultados del Algoritmo de Recocido Simulado



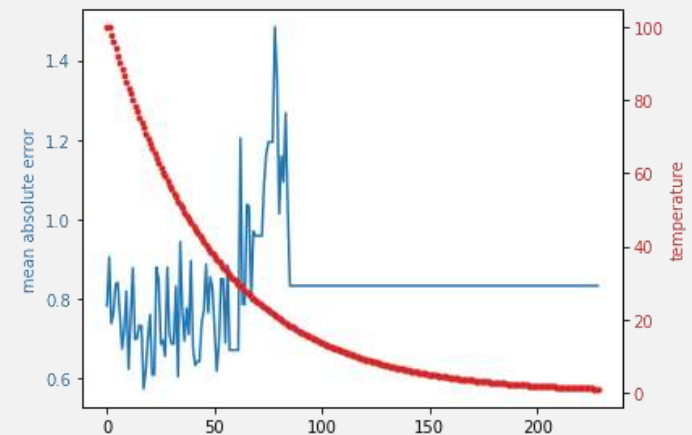
Genérico



Decreciente



Genérico sin tiempo

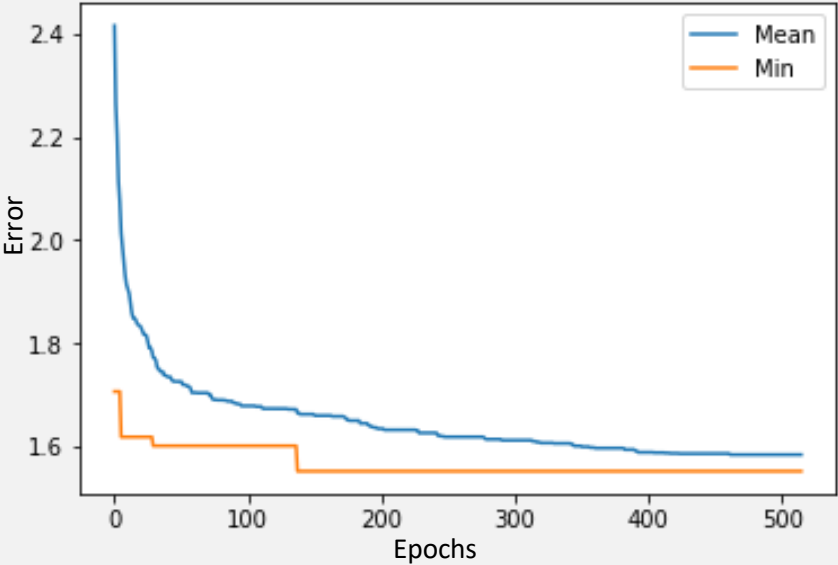


Decreciente sin tiempo

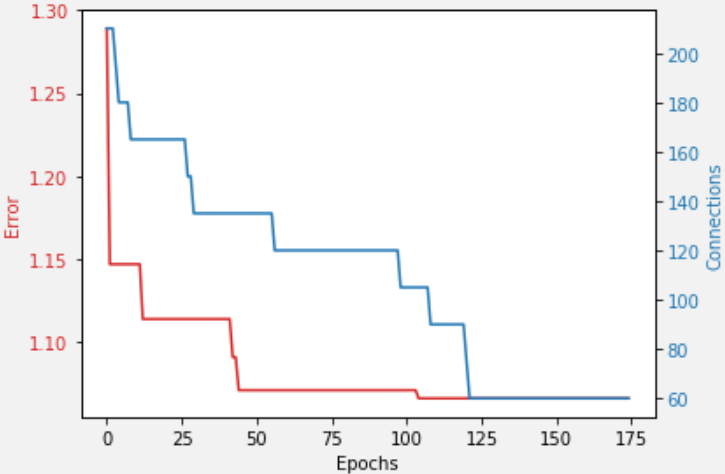
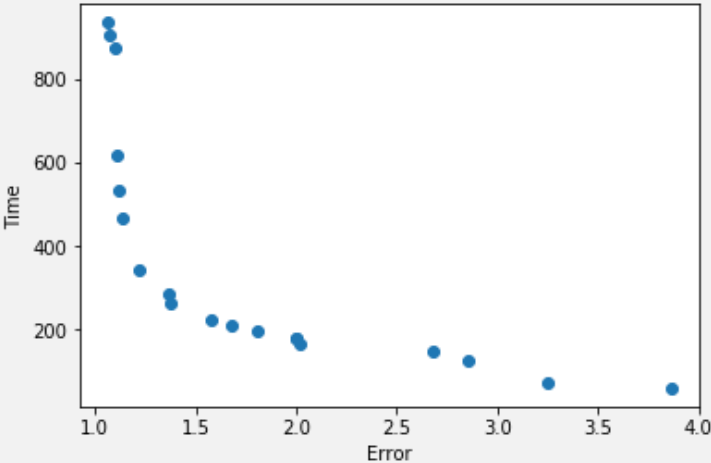
IDEA PRINCIPAL	CONCEPTOS	IMPLEMENTACIÓN	CONCLUSIONES
Hipótesis FBA	Hipótesis optimización		Solución final

Resultados de los Algoritmos genéticos

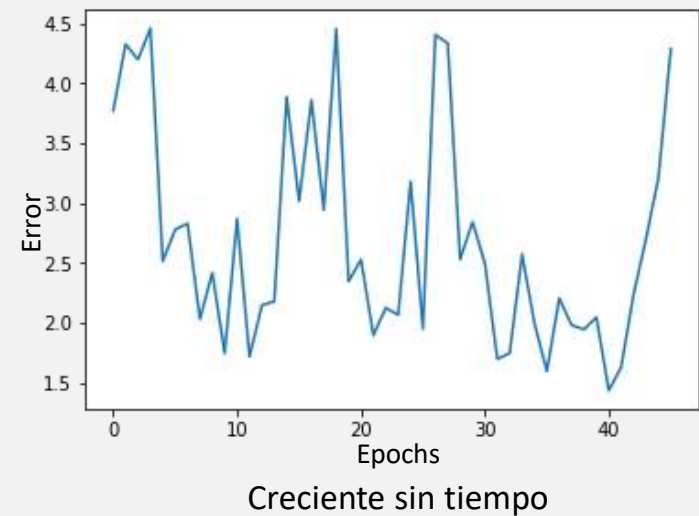
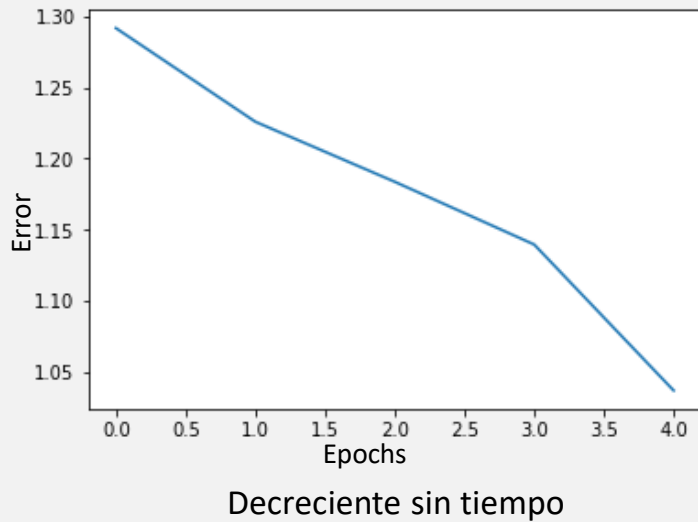
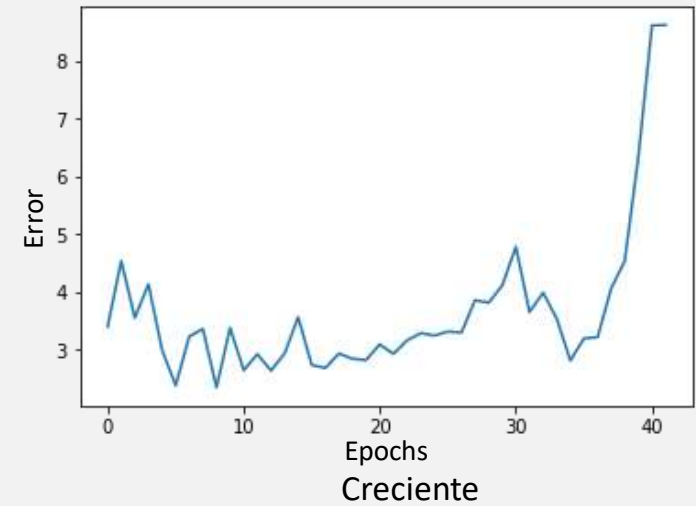
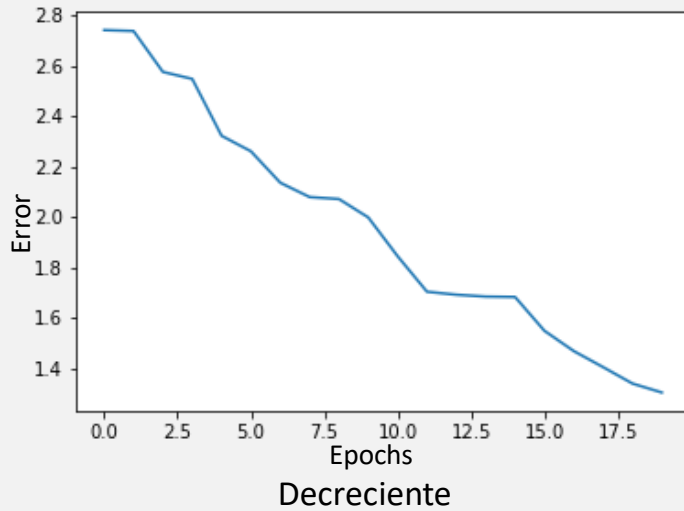
Uniobjetivo



Multiobjetivo

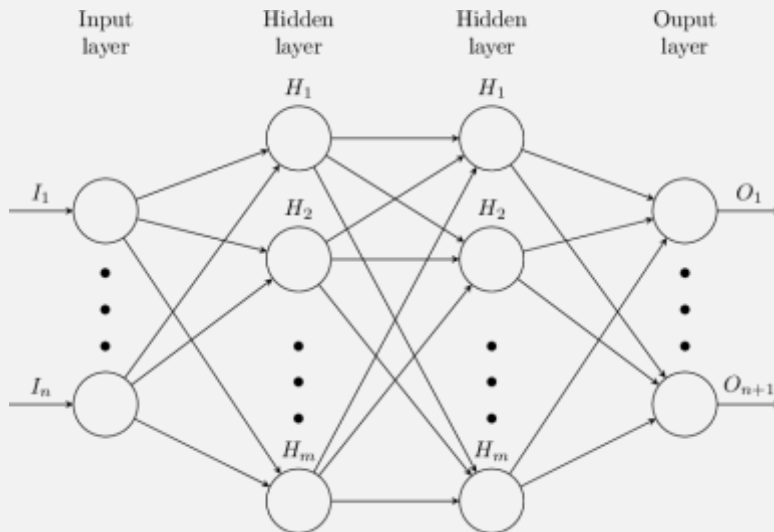


Resultados de Métodos Exhaustivos

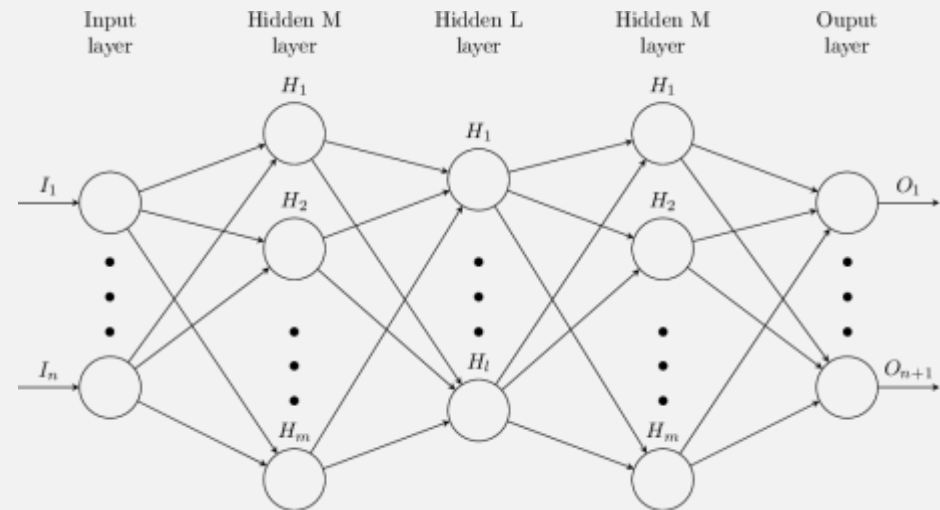


Búsqueda exhaustiva sobre estructura prediseñada

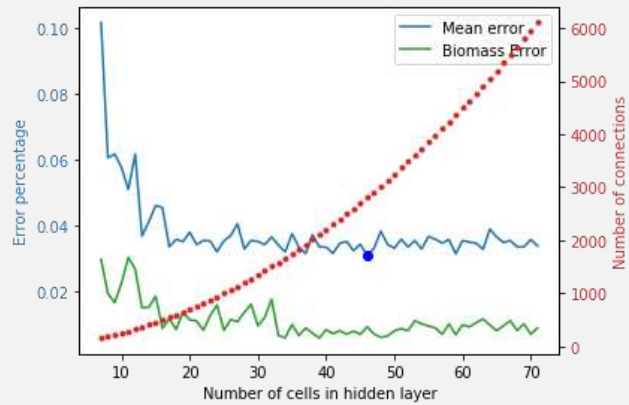
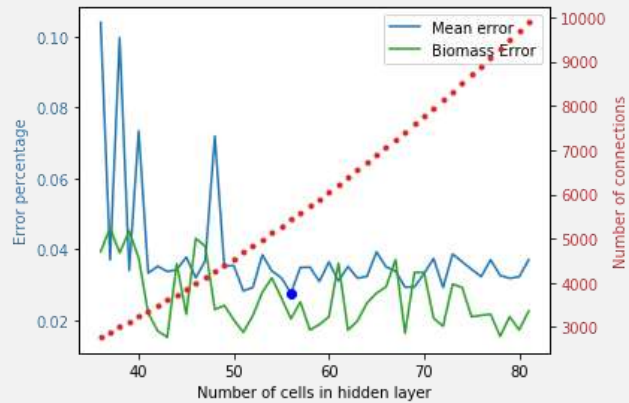
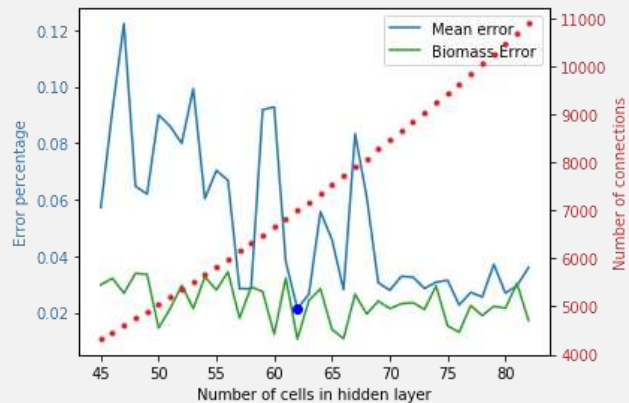
Diseño 2 capas



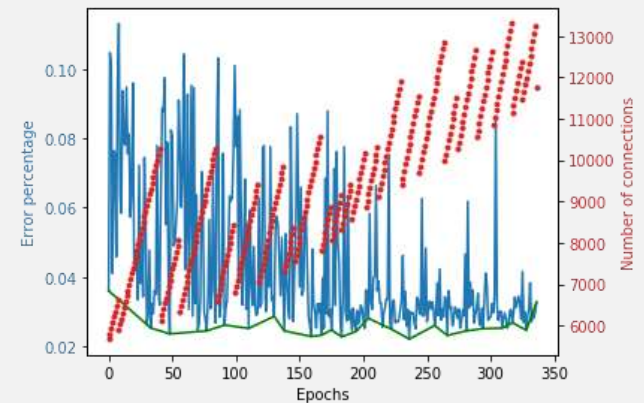
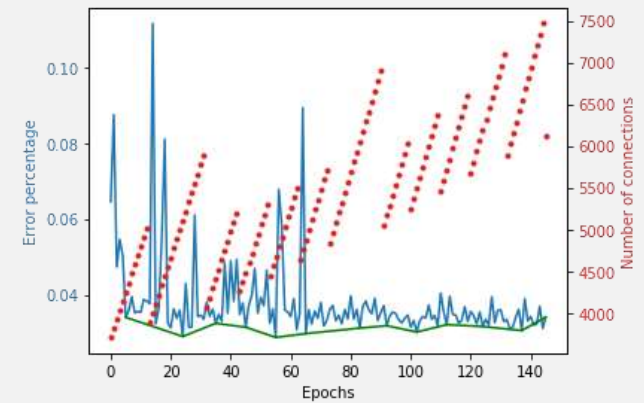
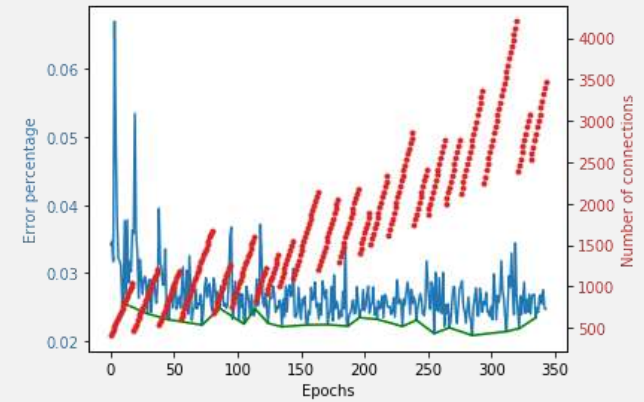
Diseño 3 capas



Diseño 2 capas

*E. coli Core**Salmonella**E. coli*

Diseño 3 capas



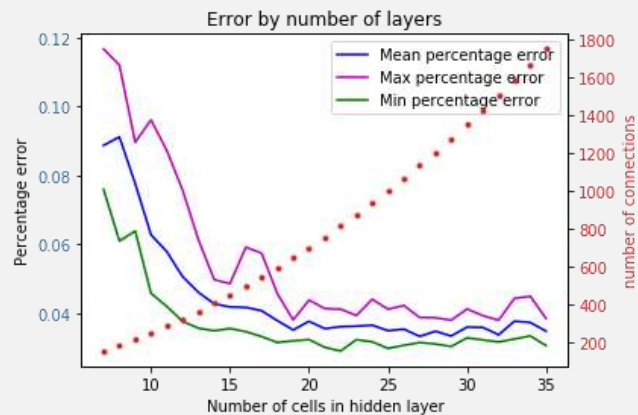
SOLUCIÓN FINAL

H1: FBA es aproximable con precisión mediante una red neural

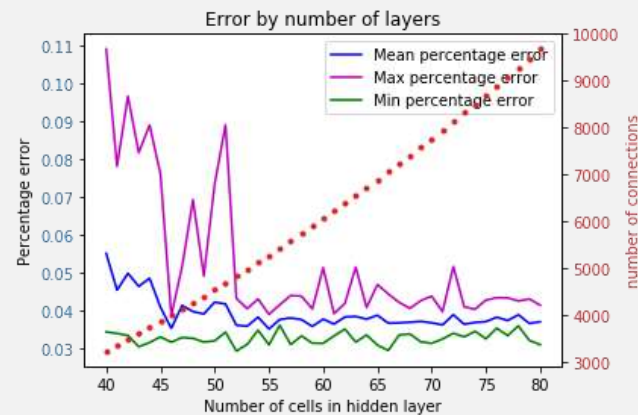
H2: La red que aproxime FBA deberá ser ejecutada rápidamente

H3: La red debe ser encontrada en un tiempo límite

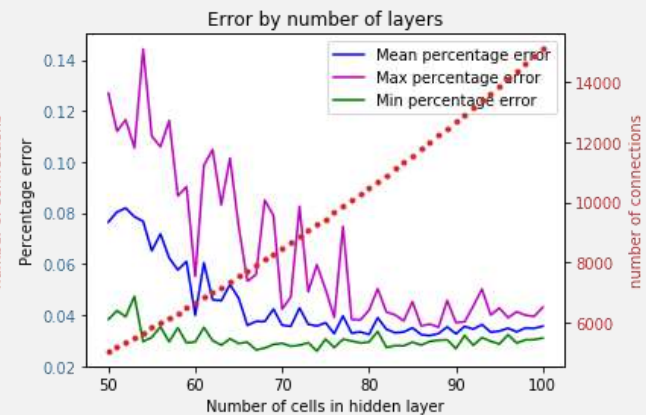
Modelos metabólicos	Error máximo	Veces más rápido que FBA	Tiempo máximo (s)
<i>E. coli Core</i>	4.1%	52.18	60
<i>E. Coli</i>	7.7%	248.36	115
<i>Salmonella</i>	5%	363.86	110



E. coli Core



Salmonella



E. coli

CONCLUSIONES Y TRABAJO FUTURO



CONCLUSIONES

- Hemos obtenido una solución que **cumple** con los **requerimientos**
- La solución obtenida es **escalable** a cualquier tipo de dato
- **No** se ha encontrado una **metaheurística** que cumpla las necesidades

TRABAJO FUTURO

- Implementación en GRO
- Mejora de la red neuronal solución
- Probar nuevos métodos de optimización
- Escalar el modelo a otros tipos de dato

Trabajo de fin de máster:

Acceleration of a bacterial metabolic simulation using neural networks with optimization techniques

Máster en Inteligencia Artificial – Universidad Politécnica de Madrid

Autor:

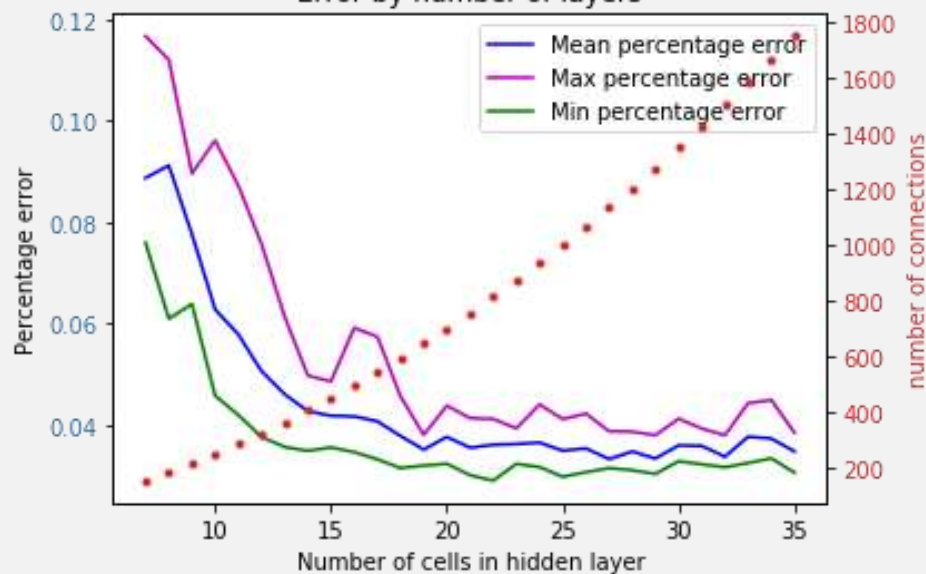
Javier París Uhryn

Tutor:

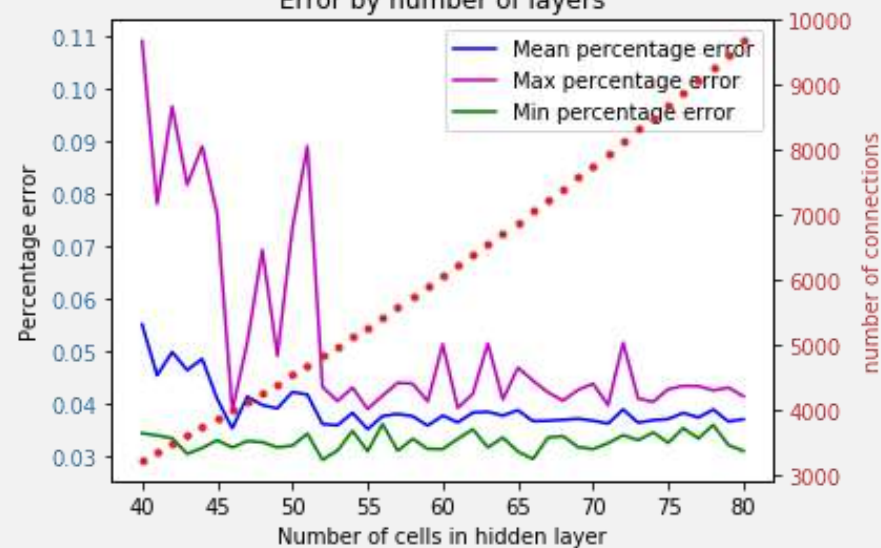
Alfonso Rodríguez-Patón Aradas

Dataset	Min proportion	Max Proportion	mean accuracy	mean connections	mean time (s)
<i>Core</i>	2	3	0.971 \pm 0.012	541.9 \pm 107.0	30.65 \pm 5.25
	3	4	0.971 \pm 0.003	931.0 \pm 122.3	52.04 \pm 7.40
	4	5	0.972 \pm 0.002	1434.8 \pm 156.7	86.84 \pm 8.60
<i>Salmonella</i>	2	3	0.965 \pm 0.015	4363.1 \pm 823.8	49.12 \pm 15.01
	3	4	0.970 \pm 0.003	7540.3 \pm 1075.9	97.38 \pm 12.74
	4	5	0.969 \pm 0.002	11696.4 \pm 1242.3	158.98 \pm 19.03
<i>Ecoli</i>	2	3	0.946 \pm 0.023	6936.7 \pm 1324.5	49.68 \pm 13.40
	3	4	0.966 \pm 0.004	12243.3 \pm 1516.5	97.88 \pm 16.82
	4	5	0.965 \pm 0.003	18276.6 \pm 2063.2	158.12 \pm 17.46

Error by number of layers



Error by number of layers



Error by number of layers

