



# RECONOCIMIENTO DEL LENGUAJE CON UNA RED NEURONAL CONVOLUCIONAL

Benjamín Galindo de la Rosa, Juan Castillo Mixcóatl, Severino Muñoz Aguirre, Georgina Beltrán Pérez  
Facultad de Ciencias Físico Matemáticas



Optoelectrónica y Fotónica

## Resumen

Hoy en día, la inteligencia artificial juega un papel muy activo en diferentes actividades humanas. Por ejemplo, es muy común tener algoritmos de inteligencia artificial en la identificación de personas en una fotografía en las redes sociales, al igual que la lista de sugerencias en servicios de video y audio en streaming, entre otros muchos ejemplos. Los algoritmos detrás de estos programas tienen una cualidad común que es emular la inteligencia humana, en el sentido de adquirir conocimiento por sí mismos para posteriormente tomar decisiones. Actualmente existen distintos algoritmos que forman parte del llamado aprendizaje de maquina o aprendizaje profundo los cuales se inspiran en las redes neuronales biológicas, de aquí el nombre de redes neuronales. En este trabajo se presentan los resultados experimentales de una red neuronal entrenada para reconocer el lenguaje hablado, particularmente la pronunciación de vocales. La red neuronal que se propuso consta de una capa de filtros convolucionales como extractor de características para lograr la identificación.

### ¿Qué es el aprendizaje automático?

Se dice que un programa computacional aprende de la experiencia E respecto de alguna tarea T y alguna medida de desempeño P, si su desempeño sobre T, al ser medido por P, mejora con la experiencia E

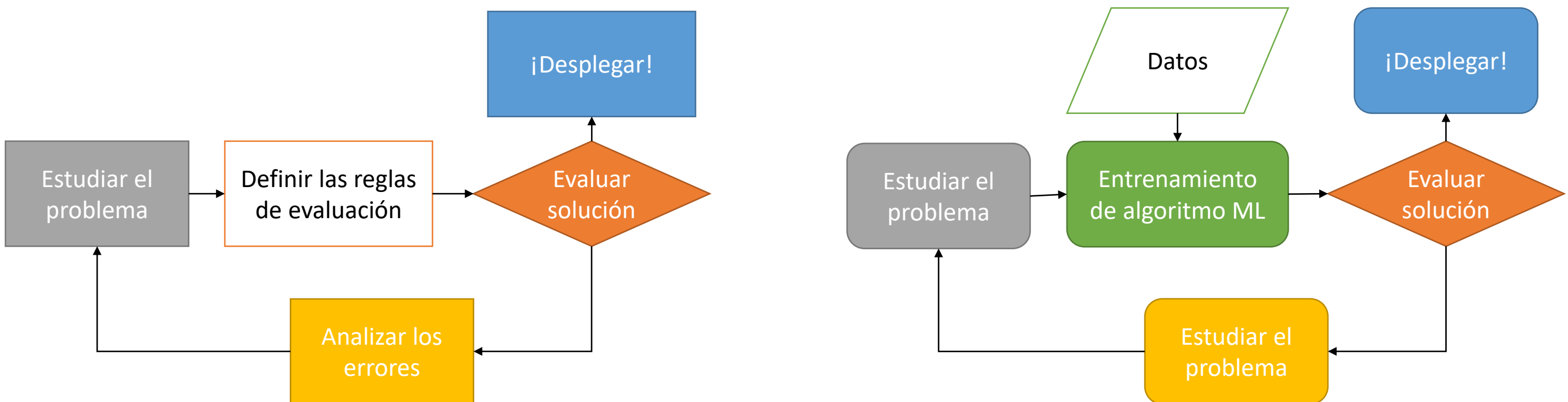


Figura 1: El enfoque tradicional

Figura 2: El método del Aprendizaje Automático

### La Neurona y las Redes Neuronales Artificiales

Las neuronas del SNC de los mamíferos tienen formas y diversos tamaños. La mayoría tiene las mismas partes que la neurona motora espinal típica mostrada en la **Figura 3**. El cuerpo celular (*soma*) contiene el núcleo y es el centro metabólico de la neurona. Las neuronas tienen varias prolongaciones llamadas *dendritas* que se extienden fuera del cuerpo celular tendiendo a ramificarse. Una neurona típica también cuenta con un largo *axón* fibroso que se origina en un área algo engrosada del cuerpo celular llamada cresta axónica. El axón se divide en terminaciones presinápticas, cada una de las cuales termina en varios *botones sinápticos* también llamados *botones terminales*. Contienen gránulos o vesículas en los que se almacenan los transmisores sinápticos que secretan los nervios. Según el número de proyecciones que surjan del cuerpo celular, las neuronas pueden clasificarse como *unipolares*, *bipolares* o *multipolares*.

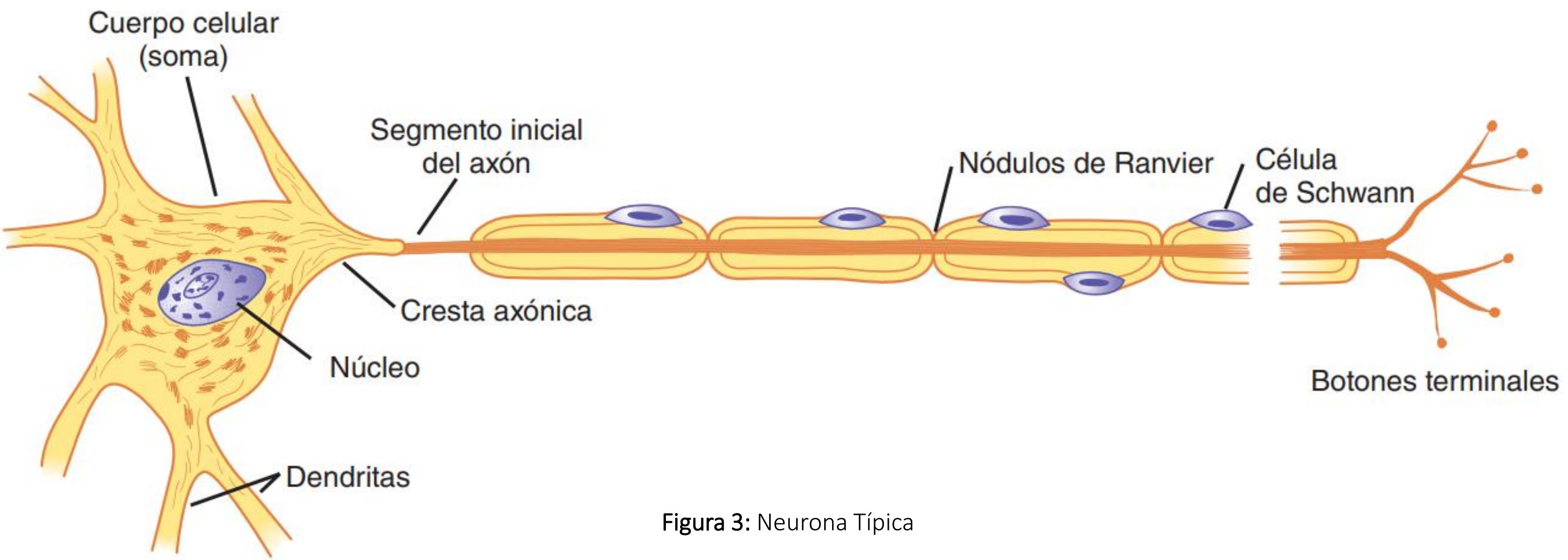


Figura 3: Neurona Típica

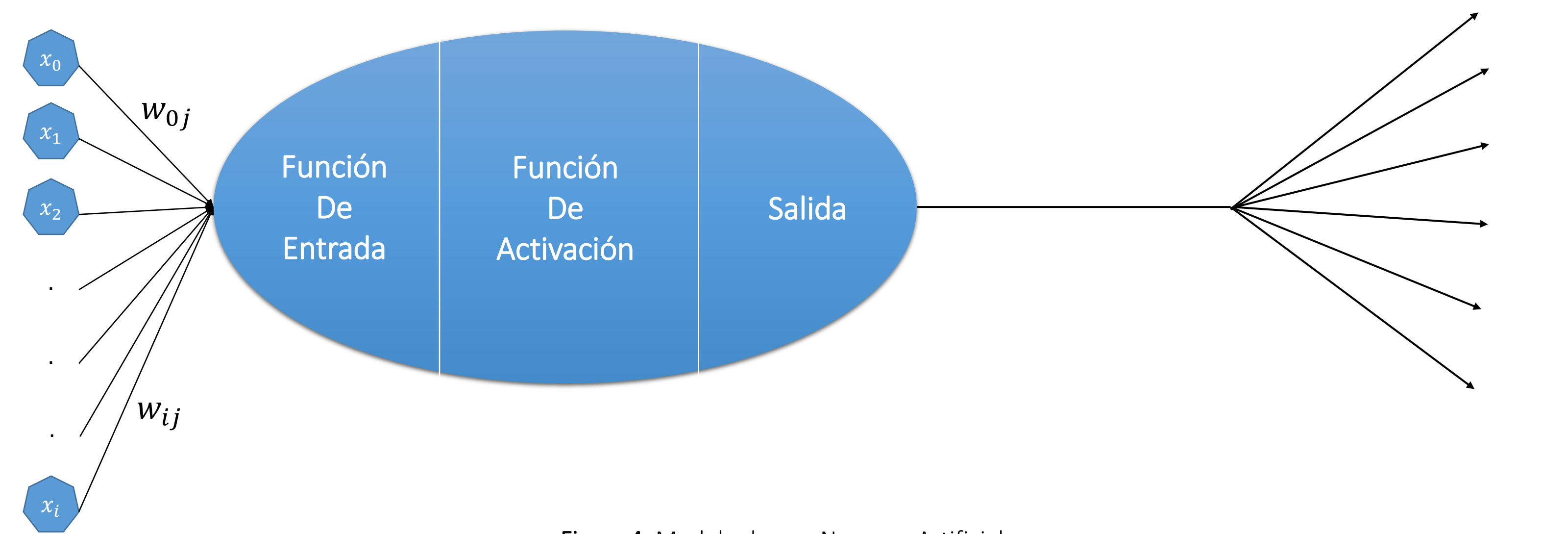


Figura 4: Modelo de una Neurona Artificial

La **Figura 4** muestra un modelo matemático simple de la neurona. A diferencia del comportamiento electroquímico de su contraparte primaria, esta se “dispara” cuando una combinación lineal de sus entradas excede un umbral predeterminado, es decir, se implementa un clasificador lineal. Una vez dicho esto, podemos definir a una **Red Neuronal** como la colección de unidades neuronales conectadas entre sí. Las propiedades de dicha red están determinadas por su topología y las propiedades de las “neuronas” en cuestión.

$$Salida = g_{activacion}(f_{entrada}) = \sigma\left(\sum_{i=0}^n w_{ij}x_i\right)$$

Figura 5: Función de activación neuronal

### ¿Qué es una Red Neuronal Convolutiva?

Una red neuronal convolutiva es un tipo de red neuronal artificial donde cada una de las neuronas artificiales corresponde a campos receptivos de una manera muy similar a las neuronas pertenecientes a la corteza visual primaria del cerebro biológico. Este tipo de red es considerado como una variación de un perceptrón multicapa, pero debido a que su aplicación es realizada en matrices bidimensionales, son muy efectivas para tareas de visión artificial como en la clasificación y segmentación de imágenes entre otras aplicaciones.

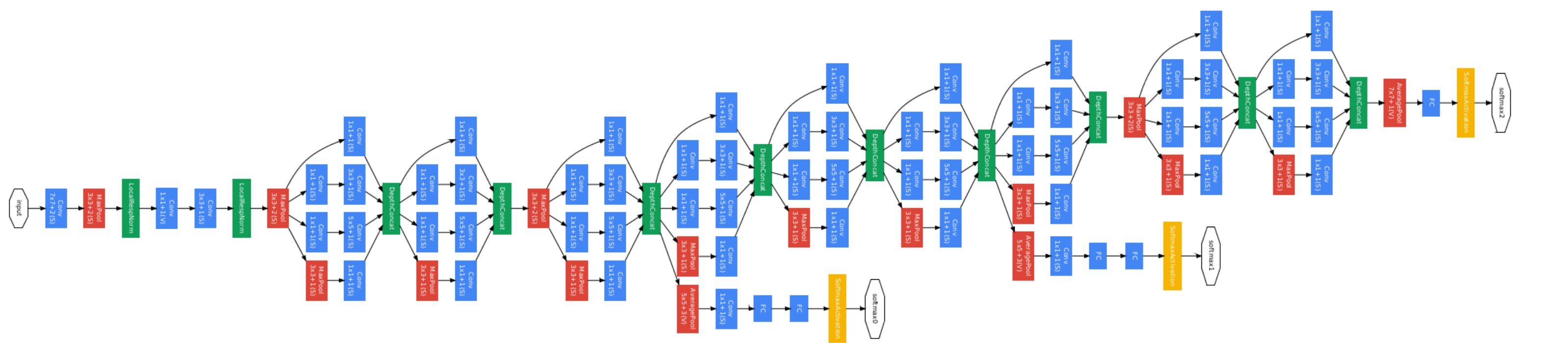


Figura 6: GoogLeNet CNN

### El modelo de regresión Softmax

El modelo de regresión logística puede ser generalizado para mantener múltiples clases directamente sin tener que entrenar ni combinar múltiples clasificadores binarios. Este procedimiento lleva por nombre Regresión Softmax, también conocida como Regresión Logística Multinomial.

La idea es simple: Dado un dato de entrada  $\mathbf{x}$ , el Modelo de Regresión Softmax calcula primeramente una puntuación  $s_k(\mathbf{x})$  para cada clase y en seguida se estima la probabilidad de cada clase al aplicar la función Softmax (también llamada Exponencial Normalizada) a cada puntuación.

$$s_k^{(i)} = \mathbf{w}^{(k)T} \mathbf{x}^{(i)}$$

Puntuación Softmax para la clase k (combinación lineal)

Note que cada clase cuenta con su propio vector de parámetros  $\mathbf{w}^{(k)}$ . Todos estos vectores se encuentran almacenados como filas en la matriz de parámetros  $\mathbf{W}$ . Una vez calculada la puntuación de cada clase para el dato de entrada  $\mathbf{x}$ , se puede estimar la probabilidad  $\hat{p}_k$  de que el dato pertenece a la clase  $k$  al evaluar la puntuación obtenida en la función Softmax.

$$\sigma\left(s_k^{(i)}\right)=\frac{e^{s_k}}{\sum_j^K\left(e^{s_j}\right)}=\frac{e^{s_k}}{e^{s_1}+e^{s_2}+\dots+e^{s_K}}$$

Función Softmax (Función de Activación)

Ahora que sabemos como es que el modelo estima probabilidades y realiza predicciones, es hora de enfocarnos en el entrenamiento, el objetivo es contar con un modelo que estime la probabilidad para la clase objetivo. Minimizar la función de costo llamada **Entropía cruzada** nos acerca a nuestro objetivo debido a que penaliza a nuestro modelo cuando la probabilidad predicha difiere de la probabilidad correcta (error) de nuestra clase objetivo.

$$J(W)=-\frac{1}{m} \sum_{i=1}^m\left[\sum_{k=1}^K\left(y_k^{(i)} \log \left(\sigma\left(s_k^{(i)}\right)\right)\right)\right]$$

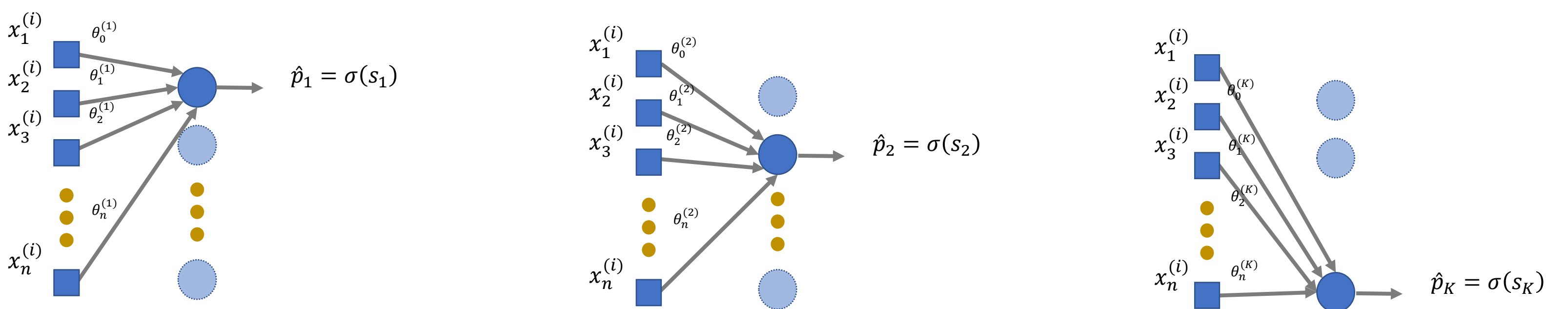
Función de Costo (Entropía cruzada)

El vector gradiente de esta función de costo respecto de  $\mathbf{w}^{(k)}$  está dada por:

$$\nabla_{\mathbf{w}^{(k)}} J(W)=\frac{1}{m} \sum_{i=1}^m\left(\sigma_k^{(i)}-y_k^{(i)}\right) \mathbf{x}^{(i)}$$

Gradiente del k-ésimo vector de parámetros

Las siguientes imágenes muestran la arquitectura del modelo de regresión Softmax.



### Desarrollo y Resultados Experimentales

Es de nuestro interés desarrollar un agente enfocado en el reconocimiento del lenguaje natural. En primer lugar, se desarrolló un clasificador binario con el fin de entender el comportamiento del mismo.

Una vez realizado, proseguimos a estudiar el comportamiento del Modelo de Regresión Logística Softmax, el cual (como primer acercamiento y ejercicio didáctico) nos ayudará a diferenciar entre diversos tipos de flores Iris (Virgínica, Versicolor y Setosa). Los resultados de dicho ejercicio se muestran en la parte izquierda de este párrafo.

En segundo lugar, buscaremos desarrollar la red neuronal convolutiva necesaria en orden a poder vaciar en ella datos de audio que nos permitan entrenar dicha red y así lograr nuestro cometido.

A continuación se comparte el QR que contiene el enlace hacia el repositorio de GitHub que estará albergando el código y los avances propios de este proyecto así como la bibliografía del mismo.

