

# Informe de Resultados — Trabajo Práctico 2

## Materia: Aprendizaje Automático Avanzado

Profesores: Dr. Juan Santos, Lic. Elimiano Churruca

Alumnos: Franco Serafini, Seivane Nicolás, Matías Cisnero

## Resumen general

### Métricas Utilizadas

Para la realización del siguiente informe se utilizará la métrica *Accuracy*, la cual representa el porcentaje de predicciones correctas respecto al total de predicciones realizadas. En este caso, se introduce una ligera modificación en su aplicación. Para la predicción de la palabra siguiente, se considera el valor con mayor probabilidad obtenido por la función *soft-max* al momento de la inferencia, lo que corresponde a la métrica **Accuracy**.

$$\text{Accuracy} = \frac{\text{Número de predicciones correctas}}{\text{Número total de muestras}} \quad (1)$$

### Arquitecturas utilizadas para *one-hot*

En esta sección analizamos las arquitecturas cuya entrada son los *embeddings* obtenidos en el trabajo práctico anterior, mientras que la salida esperada es un vector *one-hot* de dimensión igual al tamaño del vocabulario. En dicho vector, el índice con mayor valor (probabilidad) corresponde a la palabra siguiente que la red debe predecir.

Inicialmente se consideró generar los vectores *one-hot* completos para utilizarlos como salida. Sin embargo, esta estrategia produjo problemas de memoria debido al pesado tamaño de estos vectores, lo cual hacía imposible e incluso inviable el entrenamiento. Para resolver este inconveniente, se utilizó la representación *SparseCategoricalCrossentropy*, que permite indicar la clase correcta mediante un único valor entero: el índice que debería tener valor 1 en el vector *one-hot*. De esta manera, se evita la creación de vectores completos y se reduce significativamente el uso de memoria sin alterar el comportamiento del modelo.

Se evaluaron distintas arquitecturas de redes neuronales multicapa, así como diferentes valores de la tasa de aprendizaje ( $\eta$ ), en algunas casos se pudo guardar el valor de  $\eta$ , en otros casos solo reporta el valor de la métrica con el mejor  $\eta$  que se encontro.

**Arquitectura 1:** En la Tabla 1 se encuentra la primera arquitectura.

Capa	Neuronas	Activación
Entrada ( <i>embeddings</i> )	(1300)	–
Densa	(1024)	gelu
Densa	(502)	gelu
Densa	(1024)	gelu
Salida	(28030)	softmax

Cuadro 1: Resumen de la arquitectura MCP con salida one-hot 1.

**Resultados finales de la arquitectura 1:** En la Tabla se encuentran los resultados.

Métrica	Valor
Accuracy	60.00 %

Cuadro 2: Resultados de la Arquitectura 1

**Arquitectura 2:** En la Tabla 3 se encuentra la segunda arquitectura.

Capa	Neuronas	Activación
Entrada ( <i>embeddings</i> )	(1300)	–
Densa	(2000)	gelu
Densa	(3000)	gelu
Densa	(5000)	gelu
Densa	(6000)	gelu
Salida	(28030)	softmax
<b>Valor de <math>\eta</math></b>	0.001	–

Cuadro 3: Resumen de la arquitectura MCP con salida one-hot 2.

**Resultados finales de la arquitectura 2:** En la Tabla 4 se encuentran los resultados.

Métrica	Valor
Accuracy	80.00 %

Cuadro 4: Resultados de la Arquitectura 2

**Arquitectura 3:** En la Tabla 5 se encuentra la tercera arquitectura.

Capa	Neuronas	Activación
Entrada ( <i>embeddings</i> )	(1300)	–
Densa	(2000)	gelu
Densa	(3000)	gelu
Densa	(5000)	gelu
Densa	(6000)	gelu
Salida	(28030)	softmax
<b>Valor de <math>\eta</math></b>	0.004	

Cuadro 5: Resumen de la arquitectura MCP con salida one-hot 3.

**Resultados finales de la arquitectura 3:** En la Tabla se encuentran los resultados.

Métrica	Valor
Accuracy	90.00 %

Cuadro 6: Resultados de la Arquitectura 3

**Arquitectura 4:** En la Tabla 7 se encuentra la primera arquitectura.

Capa	Neuronas	Activación
Entrada ( <i>embeddings</i> )	(1300)	–
Densa	(1024)	gelu
Densa	(502)	gelu
Densa	(502)	gelu
Densa	(1024)	gelu
Salida	(28030)	softmax

Cuadro 7: Resumen de la arquitectura MCP con salida one-hot 4.

**Resultados finales de la arquitectura 4:** En la Tabla 8 se encuentran los resultados.

Métrica	Valor
Accuracy	60.00 %

Cuadro 8: Resultados de la Arquitectura 4

**Arquitectura 5:** En la Tabla 9 se encuentra la quinta arquitectura.

Capa	Neuronas	Activación
Entrada ( <i>embeddings</i> )	(1300)	–
Densa	(1024)	gelu
Densa	(2000)	gelu
Densa	(3000)	gelu
Densa	(7700)	gelu
Salida	(28030)	softmax

Cuadro 9: Resumen de la arquitectura MCP con salida one-hot 5.

**Resultados finales de la arquitectura 5:** En la Tabla 10 se encuentran los resultados.

Métrica	Valor
Accuracy	60.00 %

Cuadro 10: Resultados de la Arquitectura 5

## Arquitecturas utilizadas para *embeddings*

En esta sección analizamos las arquitecturas cuya entrada son los *embeddings* obtenidos en el trabajo práctico anterior, mientras que la salida esperada es un *embedding* con la misma dimensión que el correspondiente a la palabra siguiente en la secuencia. Debido a que los valores de los *embeddings* originales presentaban un rango relativamente amplio —con un máximo de 8.02 y un mínimo de -8.16—, se empleó la función de activación *tanh* en la capa de salida, ya que su rango acotado en  $[-1, 1]$  respeta la distribución de los datos en sus valores negativos y positivos.

Se probaron diversas estrategias para adaptar el rango de los *embeddings* objetivo. Entre ellas, normalizarlos dividiendo por 1 o por el valor absoluto del máximo, aplicar una transformación lineal para llevarlos explícitamente al intervalo  $[-1, 1]$  mediante  $x' = (x/\text{máx})| \times 2 - 1|$ , y también aplicar directamente la función  $\tanh(x)$ . Tras varios ensayos, la normalización más estable fue dividir los vectores por el valor absoluto máximo del conjunto, preservando así la estructura relativa de los *embeddings*.

A pesar de estas pruebas, los resultados obtenidos no fueron satisfactorios: la exactitud alcanzada durante el entrenamiento se mantuvo en torno al 70 %, sin mejoras significativas a pesar de ajustar la arquitectura, la función de activación y la tasa de aprendizaje. Esto evidencia que la predicción directa del *embedding* de la palabra siguiente es un problema considerablemente más complejo que la clasificación *one-hot*, ya que requiere aprender una representación continua de alta dimensionalidad.

Para evaluar qué tan similar era el *embedding* generado por la red con respecto al *embedding* real esperado, se empleó la métrica *cosine\_similarity* de *scikit-learn*. Luego se ordena el resultado de mayor a menor y el índice que tiene un valor más alto se asocia con la siguiente palabra.

Se evaluaron distintas arquitecturas de redes neuronales multicapa, así como diferentes valores de la tasa de aprendizaje ( $\eta$ ), en algunas casos se pudo guardar el valor de  $\eta$ , en otros casos solo reporta el valor de la métrica con el mejor  $\eta$  que se encontro.

**Arquitectura 6:** En la Tabla 11 se encuentra la sexta arquitectura.

Capa	Neuronas	Activación
Entrada ( <i>embeddings</i> )	(1300)	—
Densa	(1500)	gelu
Densa	(1000)	gelu
Densa	(700)	gelu
Densa	(250)	gelu
Salida	(28030)	tanh

Cuadro 11: Resumen de la arquitectura MCP con salida embedding 1.

**Resultados finales de la arquitectura 6:** En la Tabla se encuentran los resultados.

Métrica	Valor
Accuracy	57.00 %

Cuadro 12: Resultados de la Arquitectura 6

**Arquitectura 7:** En la Tabla 13 se encuentra la séptima arquitectura.

Capa	Neuronas	Activación
Entrada ( <i>embeddings</i> )	(1300)	–
Densa	(1500)	gelu
Densa	(1100)	gelu
Densa	(800)	gelu
Densa	(600)	gelu
Densa	(400)	gelu
Densa	(200)	gelu
Salida	(28030)	tanh
<b>Valor de <math>\eta</math></b>	0.001	–

Cuadro 13: Resumen de la arquitectura MCP con salida embedding 2.

**Resultados finales de la arquitectura 7:** En la Tabla 14 se encuentran los resultados.

Métrica	Valor
Accuracy	62.00 %

Cuadro 14: Resultados de la Arquitectura 2

**Arquitectura 8:** En la Tabla 15 se encuentra la octava arquitectura.

Capa	Neuronas	Activación
Entrada ( <i>embeddings</i> )	(1300)	–
Densa	(1024)	gelu
Densa	(512)	gelu
Densa	(512)	gelu
Densa	(300)	gelu
Salida	(28030)	tanh
<b>Valor de <math>\eta</math></b>	0.004	

Cuadro 15: Resumen de la arquitectura MCP con salida embedding 3.

**Resultados finales de la arquitectura 8:** En la Tabla se encuentran los resultados.

Métrica	Valor
Accuracy	90.00 %

Cuadro 16: Resultados de la Arquitectura 3

**Arquitectura 9:** En la Tabla 17 se encuentra la primera arquitectura.

Capa	Neuronas	Activación
Entrada ( <i>embeddings</i> )	(1300)	–
Densa	(3000)	gelu
Densa	(2500)	gelu
Densa	(1500)	gelu
Densa	(1000)	gelu
Densa	(500)	gelu
Densa	(350)	gelu
Densa	(170)	gelu
Salida	(28030)	tanh

Cuadro 17: Resumen de la arquitectura MCP con salida embedding 4.

**Resultados finales de la arquitectura 9:** En la Tabla 18 se encuentran los resultados.

Métrica	Valor
Accuracy	70.00 %

Cuadro 18: Resultados de la Arquitectura 4

## Resultados del modelo MCP con salida one-hot

A continuación se presentan ejemplos de generación de texto utilizando el modelo *Multi-Layer Perceptron* entrenado con salida *one-hot*. Este modelo alcanzó una exactitud cercana al 90 % en la tarea de predicción de la siguiente palabra.

### Ejemplo 1

- **Prompt:** encontraría a la maga? tantas veces me había bastado asomarme
- **K = 1, Longitud = 50**
- **Salida generada:**  
 encontraría a la maga? tantas veces me había bastado asomarme,  
 viniendo por la rue de seine, al arco que da al quai de conti,  
 y apenas la luz de ceniza y olivo que flota sobre el río me dejaba  
 distinguir las formas, ya su silueta delgada se inscribía en el  
 pont des arts , a veces andando de un lado

### Ejemplo 2

- **Prompt:** Esta materia es una gran
- **K = 5, Longitud = 50**
- **Salida generada:**  
 Esta materia es un gran el lugar que lo una mano más gorgoteo,  
 una librería del verso y el plata. la bohardilla de gardel apenas  
 sobre fichero y amortiguada sinfónicos da continuamente. por suerte,  
 amigos contra babs cuando los veía en la bohardilla de una mujer  
 el caravelle, y al matrimonio de que abrirlas lo ponían

**Ejemplo 3**

- **Prompt:** a pesar de todo habían podido traer una pequeña cantidad de
- **K = 1, Longitud = 50**
- **Salida generada:**  
a pesar de todo habían podido traer una pequeña cantidad de agua y algunas provisiones. quizá robadas por el soldado que sonreía sin entrar en detalles. desde luego ya no podía pasar mucho detrás en los conciertos del ataúd para hacer tiempo a poco o lo que me pasa qué hasta. no lo niegues, aparte de manera de decir

**Ejemplo 4**

- **Prompt:** fue encontrar a josiane en la galerie colbert y enterarme entre besos
- **K = 1, Longitud = 50**
- **Salida generada:**  
fue encontrar a josiane en la galerie colbert y enterarme entre besos y brincos de que ya no había laurent , que el final había festejado la sangre y el reflujo con el aire gusto de los ojos que cada vez, poner un huevo bien una tentativa. un buen rato sale en condiciones; cualquiera se la cosa rara. por eso me duele

**Resultados del modelo MCP con salida de *embeddings***

A continuación se presentan ejemplos de generación de texto obtenidos con el modelo basado en *embeddings*, cuyo objetivo era predecir directamente el vector de *embedding* de la palabra siguiente. Tal como se mencionó previamente, este enfoque alcanzó una exactitud aproximada del 70 %, sensiblemente inferior a la arquitectura con salida *one-hot*. Esto se refleja en la coherencia limitada de las secuencias generadas, donde aparecen repeticiones, pérdida de estructura sintáctica y combinaciones de palabras sin relación semántica clara.

**Ejemplo 1**

- **Prompt:** encontraría a la maga? tantas veces me había bastado asomarme
- **K = 1, Longitud = 50**
- **Salida generada:**  
encontraría a la maga? tantas veces me había bastado asomarme no estuviera dejaba con que se hierro una su? andando comiendo su olivo agua encuentro? estuviera inscribía abajo tubo encuentro no al encuentro andando agua ella no ya ella de acercarme fritas de un a encuentro? citas viejos apenas pont los estuviera una vidas fritas. nuestras charlando conti

**Ejemplo 2**

- **Prompt:** Esta materia es una gran

- **K = 5, Longitud = 50**

- **Salida generada:**

Esta materia es una gran como al piel da calle gente, con da dentífrico no como fritas con su olivo convencida yo estuviera asomaría yo yo en tubo? en en a estuviera al distinguir con estaría tan pero un conti bastado nuestras luz. era entrar calle ahora los gente sin agua que

### Ejemplo 3

- **Prompt:** a pesar de todo habían podido traer una pequeña cantidad de

- **K = 1, Longitud = 50**

- **Salida generada:**

a pesar de todo habían podido traer una pequeña cantidad de vendedora acercarme viejos translúcida dentífrico no convencida yo silueta fritas ceniza no? ella lo calle no a la gente, se papas es viejos detenida? de que quai? pont aprieta encuentro fritas no no yo encuentro fritas abajo ella pont estuviera. el no sorpresa a la bastado veces?

### Ejemplo 4

- **Prompt:** fue encontrar a josiane en la galerie colbert y enterarme entre besos

- **K = 1, Longitud = 50**

- **Salida generada:**

fue encontrar a josiane en la galerie colbert y enterarme entre besos acercarme como vidas tubo ella no fritas dentífrico encuentro no fritas no yo de su las precisas asomarme otro era ya olivo distinguir? de un piel cara su agua estuviera.

y veces seine delgada veces que la? no dejaba seine ella pont por quai? papas andando inscribía fritas

Como puede observarse, el modelo basado en *embeddings* tiende a producir secuencias con menor coherencia global y mayor presencia de palabras aisladas o repetidas. Esto confirma la dificultad del enfoque de regresión directa sobre un espacio continuo de alta dimensionalidad, así como la superioridad del enfoque *one-hot* en términos de estabilidad y desempeño.