Trabajo práctico 2

Juan José Cordero Gómez
Escuela de computación
Instituto Tecnológico de Costa Rica

Luis Diego Hidalgo Blanco
Escuela de computación
Instituto Tecnológico de Costa Rica

Ricardo Sánchez Alpizar
Escuela de computación
Instituto Tecnológico de Costa Rica

1. (100 PUNTOS) RED LSTM

b) 10 corridas para D=20, D=100 utilizando el segundo enfoque de preprocesamiento, resultados en tablas con medias y desviaciones estandar:

TABLE I
RESUMEN DE RESULTADOS PARA PREPROCESO #2

	D = 20			D = 100		
Epoch	Loss	Accuracy		Loss	Accuracy	
		Mean	Std		Mean	Std
1	16,8750248	0,953790938	0,209937574	68,91210175	0,938537461	0,240176801
2	6,426467896	0,959174518	0,19788573	7,869880676	0,951996411	0,213773816
3	4,920195103	0,959174518	0,19788573	4,832316399	0,96769852	0,176799589
4	3,686125278	0,956931359	0,203011657	3,433982134	0,967249888	0,17798186
5	3,038375854	0,961417676	0,192597321	2,911142349	0,972633468	0,163149027
6	2,346591711	0,957379991	0,201998871	2,786626339	0,966801256	0,179155204
7	2,394618273	0,964109466	0,186017213	2,344610929	0,969044415	0,173197394
8	1,776928663	0,965903993	0,18147581	2,327126741	0,969044415	0,173197394
9	1,929798841	0,962763571	0,189340638	2,422681332	0,968595783	0,174407547
10	1,638058305	0,963212203	0,188240419	2,006206989	0,969941678	0,170747824
Mean		0,963212203			0,969941678	
Std		0,188240419			0,170747824	

Observaciones: Como se puede observar, cuando $D = 20$, el loss y precisión inicial son menores comparado con $D = 100$; al aumentar la dimensionalidad de la capa de embedding, se aumenta el vector de características, esto a largo plazo puede facilitar la detección de patrones más extensos pero afecta el proceso de entrenamiento en velocidad y cantidad de épocas necesarias para llegar a ese punto. Al final de las ejecuciones, sin embargo, en ambos lados se puede ver una precisión muy similar en media 0.9632 vs 0.9699 y desviación estándar 0.1882 vs 0.1707, aunque para $D = 100$, el loss no disminuyó tanto, se mantuvo superior al inicio y al final.

c) 10 corridas para D=20, D=100 utilizando el primer enfoque de preprocesamiento, resultados en tablas con medias y desviaciones estandar:

TABLE II
RESUMEN DE RESULTADOS PARA PREPROCESO #1

	D = 20			D = 100		
Epoch	Loss	Accuracy		Loss	Accuracy	
		Mean	Std		Mean	Std
1	15,56038952	0,956034096	0,205019275	33,43926239	0,946164199	0,225693392
2	7,256878376	0,963212203	0,188240419	8	0,960520413	0,194733021
3	5,014931679	0,962314939	0,190433444	5,922861576	0,964109466	0,186017213
4	4,117300034	0,966352624	0,180319798	4,44744873	0,963212203	0,188240419
5	3,523610115	0,965455361	0,182623402	3,727796793	0,965006729	0,183762732
6	2,524313688	0,966352624	0,180319798	3,369677305	0,956034096	0,205019275
7	2,14724946	0,650067295	0,183762732	3,12027812	0,962314939	0,190433444
8	2,26207304	0,964558098	0,184893953	2,284707308	0,960520413	0,194733021
9	2,22116065	0,966352624	0,180319798	2,618330479	0,959623149	0,19684146
10	1,460817218	0,965903993	0,18147581	3,423551083	0,963212203	0,188240419
Mean		0,965903993			0,963212203	
Std		0,188240419			0,188240419	

Observaciones: Ahora bien, para las 10 corridas con $D = 20$ y $D = 100$ se observa un comportamiento similar al presentado en el punto *B*, pero menos acentuado, por lo que vamos a ver a continuación. Básicamente, y tomando como principal factor de observación la pérdida de cada uno, se evidencia que tener un valor para D mayor, puede ser más difícil de manejar que un valor bajo. Comparando lo obtenido, el valor 15.5604 obtenido para $D = 20$ contrasta con el valor 33.4393 obtenido para $D = 100$. Considerando el hecho de que estos valores altos para D son mas valiosos a largo plazo (es decir, con más *epochs*), se deja ver que en ejecuciones más cortas como las realizadas en este TP, la ejecución con más capas es la que se verá más afectada o desfavorecida. Sin embargo, algo que no se puede observar en esta ejecución, es el hecho de que a largo plazo la ejecución con $D = 100$, aunque vaya a ser más complicada y lenta a nivel de tiempo y recursos, terminará por obtener mejores valores para la precisión de predicción.

Esto último que se comenta si es posible observarlo medianamente a nivel de *accuracy*, ya que, aún teniendo un balance para la pérdida de 1.4608 contra 3.4236 (que marca mejores valores a favor del D más bajo), se encuentra que para la media y desviación estándar del $D = 100$ se obtienen valores de precisión muy similares e incluso mayores en ocasiones a los provistos por $D = 20$. Esto podría dar indicios de que a mayor cantidad de *epochs* el modelo entrenado para con $D = 100$ va a tener mayor efectividad y precisión en sus predicciones, en comparación con un modelo entrenado con $D = 20$ que, aunque sea más eficiente al principio (hablando de los valores en *epochs* tempranos de la tabla anterior), se va a ver desplazado por el modelo con más capas.

2. (30 PUNTOS EXTRA) PERCEPTRÓN MULTI-CAPA

2.1. Función "extract features dataset"

Se emplea la función solicitada durante la implementación los problemas encontrados fueron específicamente el modelo word2vec, ya que según lo solicitado el modelo es parte de los parámetros de la función, no obstante, se realiza un ajuste y se crea primero el modelo word2vec y posteriormente se aplica la función "extract features dataset" con la cantidad de features o "D", solicitada.

A continuación de se muestran los pantallazos solicitados de la creación y entrenamiento del modelo.

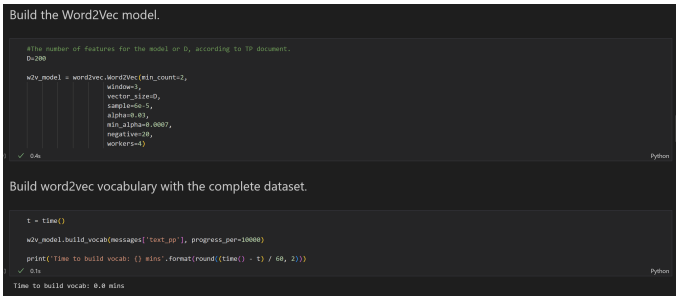

Fig. 1. Creación del modelo word2vec.

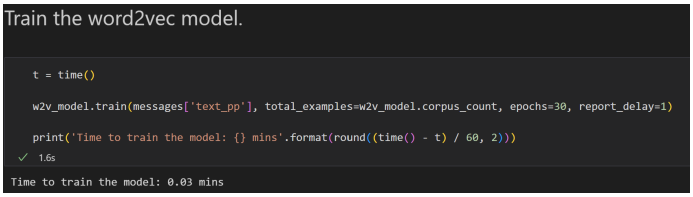

Fig. 2. Entrenamiento del modelo word2vec.

2.2. Red neuronal.

Se procede a crear un modelo de 3 capas de un perceptrón multicapa, la función de activación escogida para la salida corresponde a Sigmoid, ya que es la función que mejores resultados presentó durante las pruebas. Dentro de las capas ocultas se emplea ReLu como función de activación, ya es es la función que comúnmente genera mejores resultados. Lo anterior es basado en la lectura de [1].

Para la selección de la función de pérdida se emplea cross entropy ya es lo que comúnmente se emplea para modelos word2vec [2].

Para la función de prueba se usa la librearía sklearn.metric y se calcula el F-1 Score del modelo.

Pruebas con $D = 20$ ***y*** $M = 0.7D$*::* Empleando un $D = 20$, se calcula $M = 0.7{*}20 = 14$ neuronas en las capas ocultas del modelo y se realizan 10 corridas con dichos valores tabla III:

Seguidamente se muestra el resumen de los resultados tabla IV:

Finalmente se tiene el F-1 Score del modelo tabla V:

Pruebas con $D = 100$ ***y*** $M = 0.7D$*::* Empleando un $D = 100$, se calcula $M = 0.7 * 20 = 70$ neuronas en las capas ocultas del modelo y se realizan 10 corridas con dichos valores tabla VI:

Seguidamente se muestra el resumen de los resultados tabla VII:

Finalmente se tiene el F-1 Score del modelo tabla VIII:

TABLE III
10 CORRIDAS DEL MODELO CON 14 NEURONAS EN LAS CAPAS OCULTAS Y 20 FEATURES.

Epoch	Training loss
0	0.466092263
1	0.450312792
2	0.45031067
3	0.450309073
4	0.450307799
5	0.450306781
6	0.450305931
7	0.450305228
8	0.450304624
9	0.450304114
STD	0.004991598

TABLE IV
RESUMEN DE RESULTADOS DE PRUEBAS DEL MODELO.

Details	#
Messages Tested	1115
True Positive Tests	978
False Positive Tests	0
False Negative Tests	137
Model Accuracy (Average)	87.26%

TABLE V
F-1 SCORE.

Label	Precision	Recall	F1-score	Support
0	0.88	1	0.93	978
1	0	0	0	137
Acurracy			0.88	1115
Macro avg.	0.44	0.5	0.47	1115
Weighted avg.	0.77	0.88	0.82	1115

REFERENCES

[1] D. Gupta, "Fundamentals of deep learning – activation functions and when to use them?" Jan 2020. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/01/fundamentals-deep-learning-activation-functions-when-to-use-them/

[2] May 2022. [Online]. Available: https://datascientest.com/es/nlp-word-embedding-word2vec-es

TABLE VI
10 CORRIDAS DEL MODELO CON 70 NEURONAS EN LAS CAPAS OCULTAS Y 100 FEATURES.

Epoch	Training loss
0	0.465365875
1	0.450326103
2	0.450326103
3	0.450326103
4	0.450326103
5	0.450326103
6	0.450326103
7	0.450326103
8	0.450326103
9	0.450326103
STD	0.004755994

TABLE VII
RESUMEN DE RESULTADOS DE PRUEBAS DEL MODELO.

Details	**#**
Messages Tested	1115
True Positive Tests	973
False Positive Tests	0
False Negative Tests	142
Model Accuracy (Average)	87.26%

TABLE VIII
F-1 SCORE.

Label	**Precision**	**Recall**	**F1-score**	**Support**
0	0.87	1	0.93	973
1	0	0	0	142
Acurracy			0.87	1115
Macro avg.	0.44	0.5	0.47	1115
Weighted avg.	0.76	0.87	0.81	1115