

Práctica 7: Redes Neuronales Recurrentes

Evelyn G. Coronel
Redes Neuronales y Aprendizaje Profundo para Visión Artificial
Instituto Balseiro

(11 de noviembre de 2020)

EJERCICIO 1

Item 1

Los datos representan la cantidad de pasajeros de aviones por mes desde Enero del 1949 hasta el mes de Diciembre del 1960. Estos datos tienen una oscilación anual y crecimiento en función del tiempo como se ven en la Fig. 1.

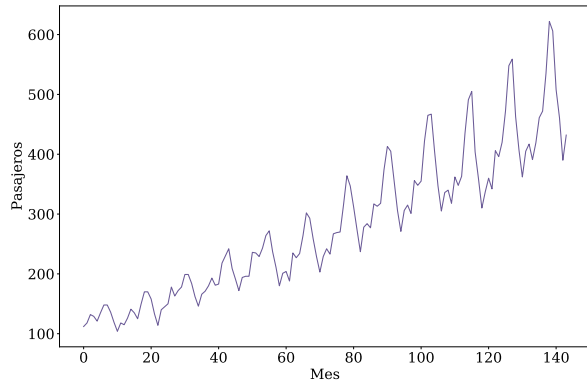


Fig. 1: Cantidad de pasajeros en función de los meses a partir de Enero del 1949

Item 2

Para utilizar estos datos para predecir la cantidad de pasajeros en función del tiempo, en el preprocesado se normalizaron los datos de tal forma que varíen entre 0 – 1.

Tomando el set de datos como $x(t)$ donde x representa la cantidad de pasajeros y t representa el mes, se organiza los datos según el desfase l que representa la cantidad de meses y se formatean según el siguiente algoritmo:

- Si $l = 1$:

$$X = x(t) \quad Y = x(t + 1)$$

- Si $l = 2$: Por ejemplo si queremos obtener Y para el primer elemento $x(0)$, obtenemos los siguientes vectores:

$$X = (x(0), x(1)) \quad Y = x(2)$$

Generalizando para cualquier t :

$$X = (x(t), x(t + 1)) \quad Y = x(t + 2)$$

- Para l arbitrario: Generalizando para cualquier t :

$$X = (x(t), x(t + 1), \dots, x(t + l - 1)) \quad Y = x(t + l)$$

Los datos se formatean de esta forma para hacer que la red aprenda a predecir la cantidad de pasajeros el mes A , utilizando la información considerando los l meses anteriores

Items 3, 4 y 5

Se agrega el ruido gaussiano para probar si la red es capaz de aprender el patrón de la cantidad de pasajeros a pesar del ruido.

Para entrenar la red recurrente, se separaron los datos de entrenamiento y test en 70 % y 30 % respectivamente.

Dado que nuestro conjunto de datos solo contiene la cantidad de pasajeros, por que al redimensionar los datos para la entrada del LSTM, se considera que cantidad de atributos (*features*) es 1 y el rango de tiempo utilizado para hacer el entrenamiento de l meses.

Item 6 y 7

Las arquitectura de la red utilizada para predecir la cantidad de pasajeros son: una capa LSTM con 4 entradas y una salida de 1, y otra red con una capa densa de 4 neuronas con una salida de 1.

Se utilizó el optimizador Adam con una tasa de aprendizaje de 2×10^{-3} y MSE como función de pérdida para ambas redes, las mismas fueron entrenadas durante 100 épocas.

Item 8 y 10

Entrenando la red con $l = 1$ se obtiene las predicciones (V) usando la capa LSTM mostrado en la Fig. 4 junto a los datos de entrenamiento (E), en caso de la red con la capa densa estas curvas se muestran en la Fig. 5. La pérdidas normalizadas de estas redes se muestran en las Figs. respectivamente, los datos de entrenamiento llegan a un valor de $MSE = 0.0023$ mientras que la test $MSE = 0.012$

con la capa de entrada del tipo LSTM, cuando pasamos a la otra red el valor de MSE para el entrenamiento es 0.0024 y para la test 0.010. En este trabajo para $l = 1$, la red con capas densas da una mejor predicción que la red LSTM.

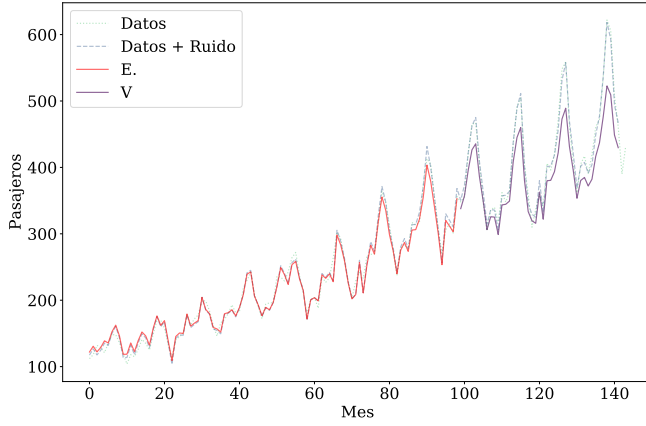


Fig. 2: Predicción de la cantidad de pasajeros usando un desfase de 1 mes para el entrenamiento usando una capa LSTM

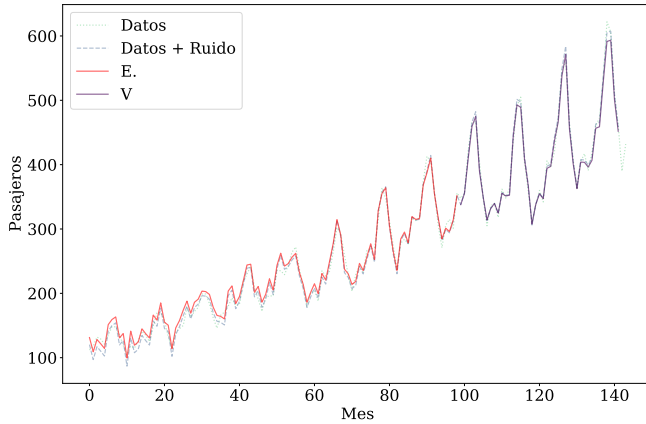


Fig. 3: Predicción de la cantidad de pasajeros usando un desfase de 1 mes para el entrenamiento usando una capa densa

Considerando que las variaciones de la cantidad de pasajeros tiene una modulación de 12 meses, este trabajo utilizó $l = 13$ para ver como funciona con red prediciendo las cantidad de pasajeros con los 13 meses anteriores. La predicción se muestra en la Fig.6 , donde se muestra que la red predice mejor los picos el caso de $l = 1$.

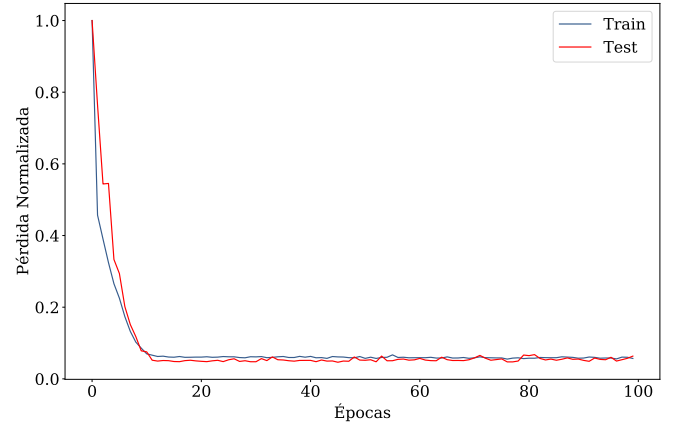


Fig. 4: Pérdida en función de las épocas usando un desfase de 1 mes para el entrenamiento usando una capa LSTM

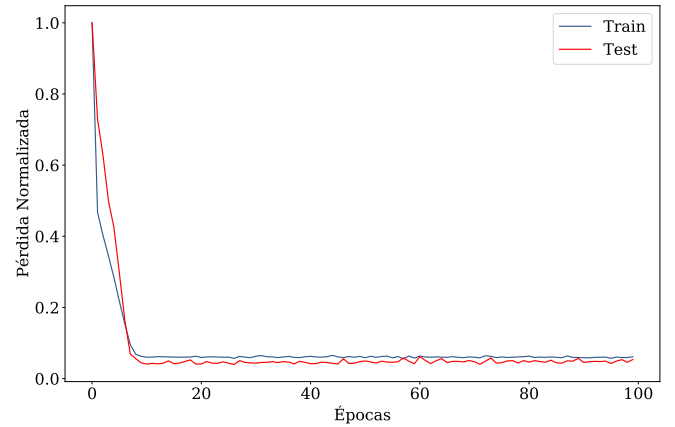


Fig. 5: Pérdida en función de las épocas usando un desfase de 1 mes para el entrenamiento usando una capa densa

Item 9

Por lo observado en los items anteriores, la predicción de la red depende de cuanto fue el desfase de tiempo dado para el entrenamiento. En la Fig. 7 se muestra el valor de MSE del test en función del parámetros l . Se observa un mínimo del valor del MSE en los desfases corridos en un año, esto se debe a la periodicidad del conjunto de datos.

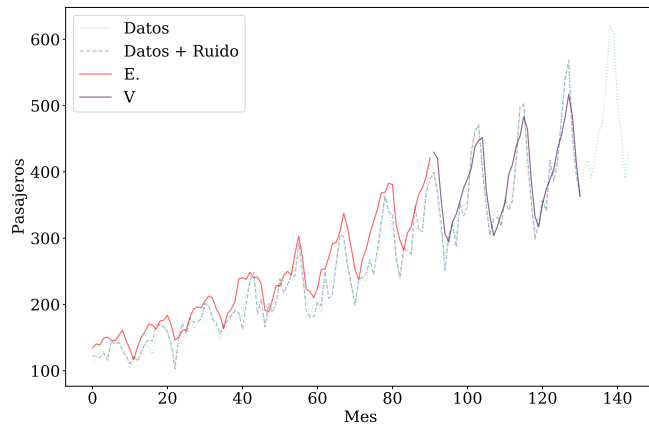


Fig. 6: Predicción de la cantidad de pasajeros usando un desfase de 13 meses para el entrenamiento usando una capa LSTM

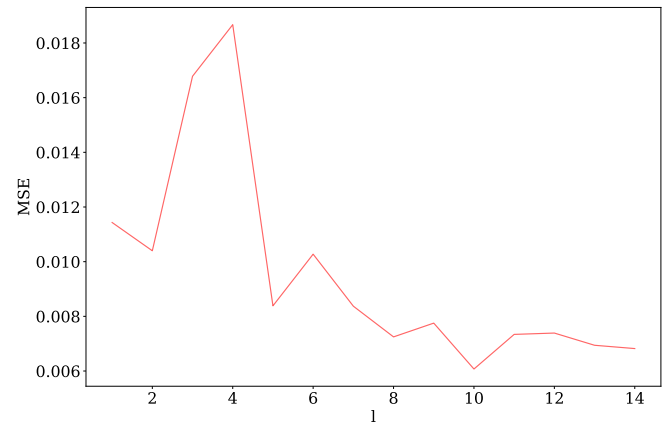


Fig. 7: Valores de MSE en función del desfase usando una capa LSTM