

Redes Neuronales Profundas

Trabajo Práctico 4

Primera parte:

Generación de texto y predicción de letras en español usando [WikiCorpus en español](#): para este TP deberá utilizar [este código](#) para leer el dataset de entrenamiento y [este ejemplo de red lstm](#) en Keras.

1. Utilice el ejemplo tal cual está (sólo adaptelo para la lectura y predicción de texto en español) para establecer un error de entrenamiento, validación y test de base, de manera de cualquier cambio debería mejorar estos valores. Responda:
 - a. ¿Cuántas iteraciones son necesarias para obtener texto que parece escrito en español?
 - b. ¿Por qué es tan costoso entrenar este tipo de modelos?
 - c. ¿Puede sugerir algún preprocesamiento para los datos?
 - d. ¿Qué efecto tiene temperatura en el sampleo de texto generado?

Segunda parte:

Combinación de los modelos de reconocimiento de letras manuscritas y predicción de texto en español: para eso es necesario integrar el código realizado en los TPs anteriores. La evaluación se realizará sobre [las imágenes de test de las palabras manuscritas del texto de Borges](#).

1. Adapte el clasificador de dígitos manuscritos para que funcione sobre una frase completa simplemente desplazando una ventana en la dirección del texto (naturalmente la predicción repetirá varias veces la misma letra)
2. Adapte el predictor de texto en español para que reciba cada letra de la clasificación del punto anterior, de manera de seleccione la letra más probable y elimine la repetición.
3. Experimente combinando las probabilidades con distintos pesos y reporte los resultados.