Índice

Prefacio **13**

Redes recurrentes. ¿Qué resuelven? **17**

1. Valor añadido de las RNN sobre las redes clásicas 18
2. Datos anidados en secuencias 24

Mecanismo de corrección LSTM **29**

1. Solución LSTM 31
2. Línea de estado (el modelo situacional) 37
3. Mecanismo de olvido 42
4. Mecanismo de suma o actualización 45
5. Mecanismo de generación de la salida de la capa oculta 50

Palabras como entradas **55**

1. Matriz de incrustación (*embeddings*) 56
2. Cómo se llega a la matriz de incrustación con word2vec 59
3. Integración e inicialización de la capa de incrustación 64
4. Actualización de la matriz de incrustación 73
5. Inicialización a partir de otra RNN (o transformer) 76
6. Inicialización a partir de funciones aleatorias 79

Topologías para modelos generativos del lenguaje **83**

1. Modelos del lenguaje (sistemas de producción) 84
2. Imagen a secuencia (uno a muchos) 98
3. Secuencia a secuencia (muchos a muchos) 101
4. Secuencia a clasificación (muchos a uno) 105

Retropropagación y Gradiente Descendente **109**

1. Introducción 110
2. Concepto de pase hacia adelante y de retropropagación 111
3. Concepto del gradiente descendente 114
4. Repaso de las funciones implicadas en una red 118

4.1. Función de integración 120
4.2. Función de activación 122
4.3. Función de evaluación (de pérdida o coste) 123
5. Actualización de los pesos y derivaciones 125
5.1. Actualización de los pesos de las conexiones que se unen a la capa de salida 125
5.2. Actualización de los pesos de las conexiones que se unen a la capa oculta 132
6. Retropropagación a través del tiempo (BPTT) 139
Hiperparámetros 147
1. Número de épocas o ciclos 148
2. Conjunto de entrenamiento, desarrollo y test 148
3. Validación cruzada (cross-validation) 149
4. Lotes y su tamaño 150
5. Inicialización de los pesos de la RNN 151
6. Inicialización de la matriz de incrustación 151
7. Actualización de la matriz de incrustación 151
8. Salvaguarda de modelos intermedios 152
9. Entrenamiento incremental 152
10. Parámetro de tasa de aprendizaje 152
11. Funciones de activación 153
11.1. Función lineal (nula o mínima modificación) 154
11.2. Función sigmoidea o logística (todo o nada) 155
11.3. Función tangente hiperbólica (fosa o cielo) 156
11.4. ReLU (unidad lineal rectificada) 157
11.5. Softmax (función exponencial normalizada) 158
12. Función de evaluación (coste o pérdida) 159
12.1. Error cuadrático medio (MSE) 160
12.2. Entropía cruzada (*Cross-entropy, CE*) 161
13. Algoritmos de optimización 162
13.1. Momento 164
13.2. RMSprop (*Root Mean Square Propagation*) 166

13.3. Adam (*Adaptive Moment Estimation*) 168
14. Olvido Catastrófico y MAS (*Memory Aware Synapse*) 170
15. Sobreajuste 176
16. Dilución (*Dropout*) 178

Extensiones de las RNN **181**
1. RNN apiladas (diversas capas de recurrencia) 181
2. RNN bidireccionales (biLSTM) 183

Mecanismo de atención y Transformers **187**
1. Capa de autoatención y el germen de los transformers 188
2. Roles de las palabras en la capa de autoatención 195
3. Autoatención bidireccional 200
4. Transformers 203
   4.1. Incrustaciones de posición 203
   4.2. Bloques transformer 205
   4.3. Bloques apilados 209
   4.4. Autoatención multicabezal 210
   4.5. Topologías 211
5. El mecanismo de atención en las RNN 218

¿Dónde está representado el conocimiento? **225**
1. Introducción a la cuestión representacional 225
2. Pesos de las propias conexiones 236
3. Pesos de la matriz de incrustación 237
4. Incrustaciones contextuales: la interacción de las entradas con los pesos. 238
5. Memoria proveniente de los estados anteriores de la capa oculta 243
6. Memoria de la línea de estado en las LSTM 244
7. Memoria del contexto en el mecanismo de atención 245

Ejemplos de codificación **247**

1. Un modelo RNN-LSTM autogenerativo con caracteres 248
2. Un modelo RNN-LSTM autogenerativo con palabras 257
3. Un modelo trasformer de clasificación de frases 261
4. Proyecto RNN-LSTM y transformer en C# (Seq2SeqSharp) 266
5. Servicios en nube (el caso de OpenAI) 268

Autoevaluación **277**

Bibliografía **311**