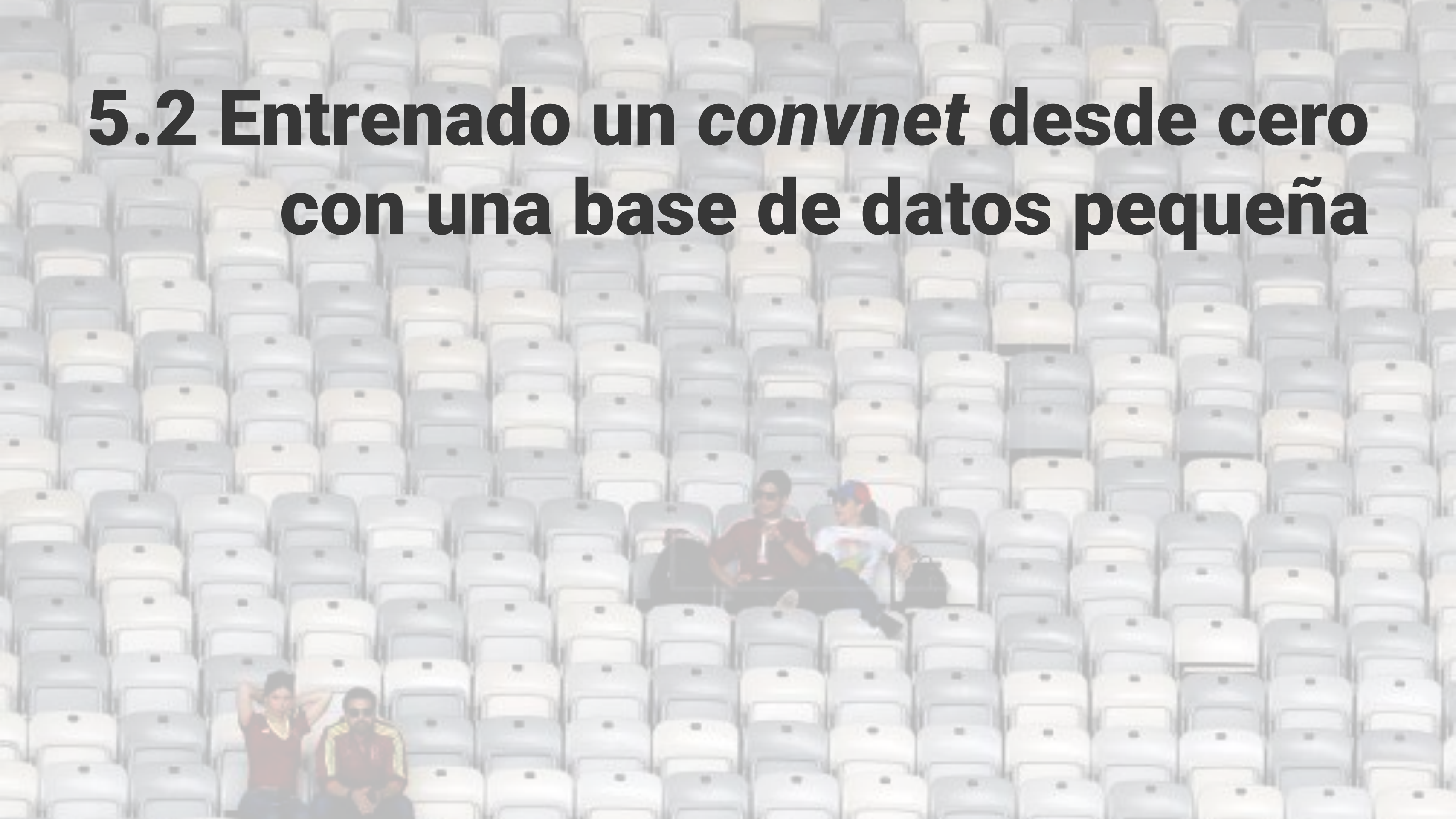


5.2 Entrenado un *convnet* desde cero con una base de datos pequeña



Base de datos pequeña

- Depende del problema
 - Pocos cientos de imágenes
 - Pocas decenas de miles de imágenes





Datos de Kaggle (2013): dogs-vs-cats¹

- Imágenes JPEG
- Resolución mediana
- Heterogéneas
 - Tamaño
 - Apariencia, ...
- 25,000 imágenes
 - 12,500 perros
 - 12,500 gatos
 - 543 MB (comprimidas)



¹ www.kaggle.com/c/dogs-vs-cats/data

Preparación

Animales	Entrenamiento	Validación	Test
 Perros	1,000	500	500
 Gatos	1,000	500	500

Selección del tipo de activación de la última capa y de la función de pérdida

Tipo de problema	Activación	Pérdida
Clasificación binaria	sigmoid	binary_crossentropy
Multiclase, 1-etiqueta	softmax	categorical_crossentropy
Multiclase, n-etiquetas	sigmoid	binary_crossentropy
Regresión a valores arbitrarios	linear2	mse
Regresión a valores entre 0-1	sigmoid	mse / binary_crossentropy

² Por defecto

Data augmentation (DA)

⚠ La ejecución es tardada

- Técnica para mitigar el sobreajuste
 - Otras técnicas: *Dropout*, *weight decay* (L1, L2 *regularization*)
- Específica al tratar con imágenes (muy utilizada)
- Sobreajuste es debido a escasez de datos
 - Un dataset infinito nunca se sobreajustaría
- DA genera más datos de entrenamiento
 - Nuevos datos a partir de la propia muestra
 - Transformaciones: tamaño | aspecto | orientación
 - Función `ImageDataGenerator`
 - La red no ve dos veces la misma imagen para entrenar
 - ⚠ Inputs correlacionados
 - Necesario para acabar de quitar sobreajuste: **dropout** | regularización | parámetros de la red

Referencias

1. Rafael Alenka 2018, *Kaggle*, [Dealing with very small datasets](#)
2. Jordi Torres, 2019, *Experfy*, [Convolutional Neural Networks for Beginners – Practical Guide with Python and Keras](#)
3. Satsawat Natakarnkitkul, 2020, *Machine Learning*, [Beginners Guide to Convolutional Neural Network from Scratch – Kuzushiji-MNIST](#)
4. Shreyak, 2020, *Becoming Human*, [Building a Convolutional Neural Network \(DNN\) Model for Image classification](#)
5. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. 2002, . *Journal of artificial intelligence research* 16: 321-57, [SMOTE: synthetic minority over-sampling technique](#)
6. Torgo L. 2017, CRC press, *Data mining with R: learning with case studies*

Convnet
preentrenado

El próximo día



KEEP
CALM
AND
SEE YOU
NEXT CLASS