

Una guía de Redes Neuronales Convolucionales (CNN)

Carro Barallobre, Icia
Álvarez Crespo, Lucía María

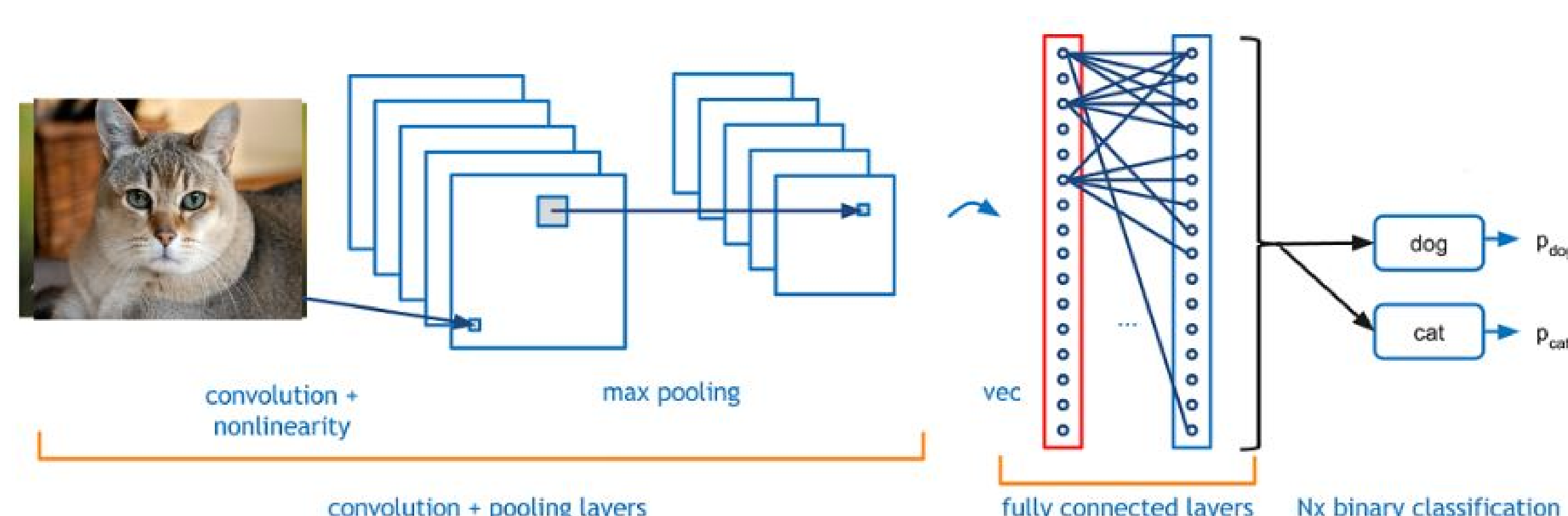
icia.carro.barallobre@udc.es

lucia.maria.alvarez.crespo@udc.es

Universidade da Coruña, 2019

CNN y su analogía con el cerebro

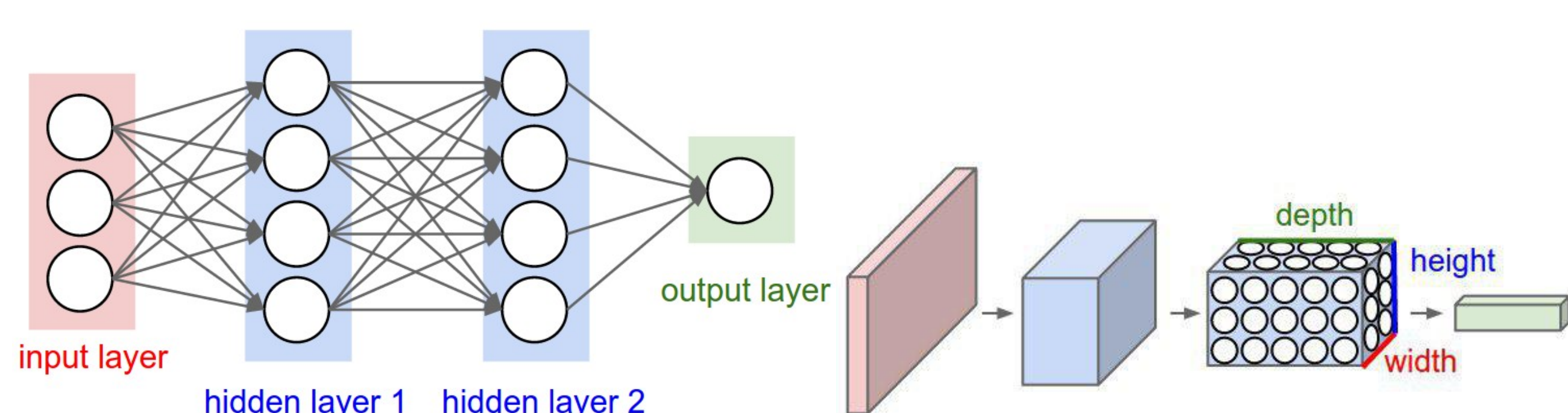
Constantemente estamos analizando el mundo que nos rodea: vemos, etiquetamos, hacemos predicciones y reconocemos patrones en función de lo que hemos aprendido. Análogamente, necesitamos mostrar imágenes a un algoritmo antes de que sea capaz de generalizar la entrada y hacer predicciones para imágenes que nunca antes había visto. Para enseñar a un algoritmo a reconocer objetos, en este caso, usaremos un tipo de red neuronal artificial (RNA): las redes neuronales convolucionales (CNN), inspiradas en el cerebro.



Arquitectura (RNA convencional vs CNN)

En las CNN las capas están organizadas en tres dimensiones: ancho, alto y profundo. Tienen dos componentes:

- La parte de extracción de características.
- La parte de clasificación: aquí las capas totalmente conectadas servirán como un clasificador por encima de estas características extraídas.



Extracción de características

Las CNN tienen la particularidad de usar aprendizaje profundo (Deep Learning) para la extracción de características; el propio algoritmo decidirá qué detalles de la imagen considera más relevantes a la hora de diferenciar los distintos objetos presentados, evitando la programación de técnicas basadas en visión artificial y con alta fiabilidad de resultados. Para entender cómo las CNN extraen características debemos entender un concepto muy importante: la convolución.

Ejemplo práctico

<https://github.com/luciamariaalvarezcrespo/poster-ipt-2019>

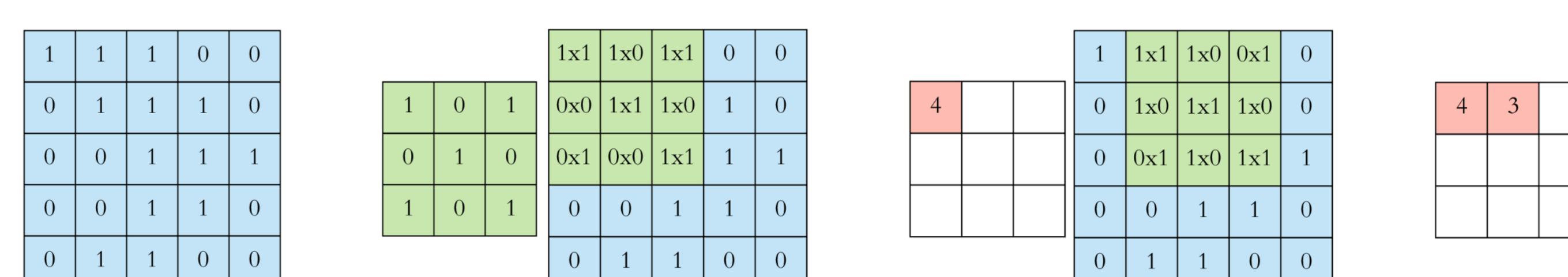
<https://github.com/IciaCarroBarallobre/poster-ipt-acm-19>

Referencias

- [1] H. Huang, H. Zhou, X. Yang, L. Zhang, L. Qi, and A.-Y. Zang. Faster r-cnn for marine organisms detection and recognition using data augmentation. *Neurocomputing*, 2019. ISSN 0925-2312. doi: <https://doi.org/10.1016/j.neucom.2019.01.084>. URL <http://www.sciencedirect.com/science/article/pii/S0925231219301274>.
- [2] W. R. Johnson, J. Alderson, D. Lloyd, and A. Mian. Predicting athlete ground reaction forces and moments from spatio-temporal driven cnn models. *IEEE Transactions on Biomedical Engineering*, 66(3):689–694, March 2019. ISSN 0018-9294. doi: 10.1109/TBME.2018.2854632.
- [3] S. Mejri, C. Franchet, R. Ismat-Ara, J. Mothe, P. Brousset, and E. Faure. Deep analysis of cnn settings for new cancer whole-slide histological images segmentation: the case of small training sets. *6th International Conference on Bioimaging*, 2019.

Convolución

La convolución es la fusión de dos conjuntos de información: en el caso de las CNN, la convolución se realiza en los datos de entrada con el uso de un filtro para producir un mapa de características. Se ejecuta una convolución deslizando el filtro sobre la entrada. Realizaremos numerosas circunvoluciones en nuestra entrada, donde cada operación utilizará un filtro diferente: esto resulta en diferentes mapas de características. Al final, tomamos todos estos mapas de características y los juntaremos como la salida final de la capa de convolución.

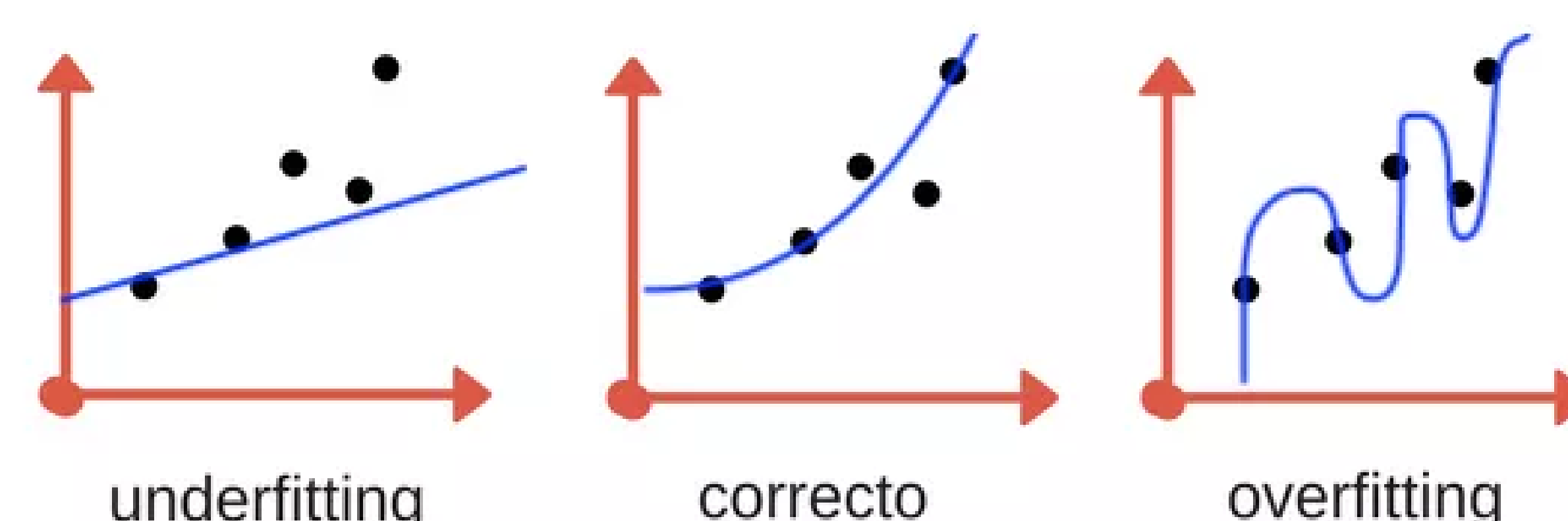


Clasificación

Después de la convolución y la agrupación de capas, la parte de clasificación consta de unas pocas capas totalmente conectadas. Sin embargo, estas capas totalmente conectadas solo pueden aceptar datos de una dimensión, por lo que debemos convertir nuestro volumen 3D en un vector 1D.

Entrenamiento

Al igual que a un niño, nuestra CNN necesita que le mostremos ejemplos para poder aprender a diferenciar objetos. Aunque es lógico pensar que cuantos menos ejemplos le demos, menos probabilidad tendrá de etiquetar bien nuevos casos y, cuantas más le demos, mejor lo hará: esto no es así. Si bien es cierto que con pocos ejemplos (underfitting) no clasificaremos correctamente, cuando un sistema se entrena demasiado o se entrena con datos extraños (overfitting), se ajusta a unas características muy específicas de los datos de entrenamiento y no es capaz de generalizar.



Conclusión

En resumen, las CNN son especialmente útiles para la clasificación y el reconocimiento de imágenes. Contienen dos partes: una parte de extracción de características y una parte de clasificación. La principal técnica especial en las CNN es la convolución: donde un filtro se desliza sobre la entrada y combina el valor de entrada más el valor del filtro en el mapa de características.

Trabajo futuro

A día de hoy las CNN ya se están usando en numerosos campos con muy buenos resultados. Algunos de ellos son el reconocimiento de organismos marinos[1], aplicaciones en física para cálculo de fuerzas y momentos de reacción[2] o la detección automática del cáncer[3].