

Tema 4.2

Regularización sobre datos

Deep Learning

Máster Oficial en Ingeniería Informática

Universidad de Sevilla

Contenido

- Preprocesamiento de datos
- Aumentado de datos

Preprocesamiento de datos

- Es importante **normalizar** los datos para el descenso por gradiente

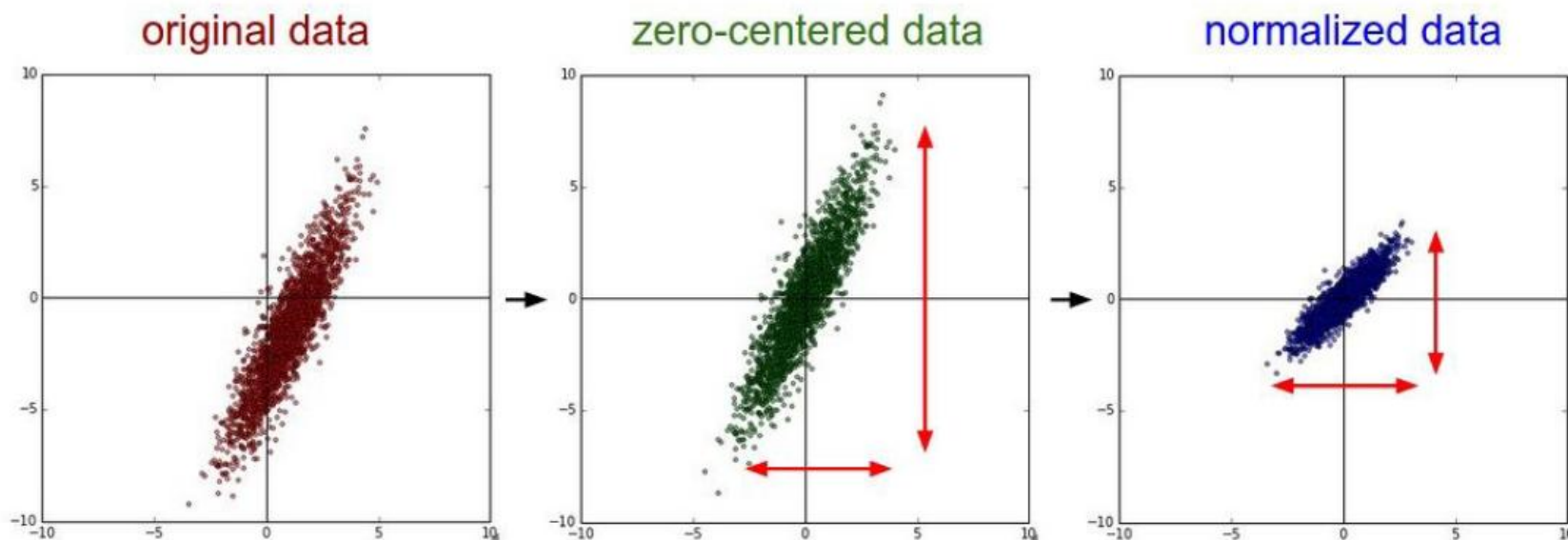
- La actualización del peso j depende de la entrada x_j :

$$\theta_j = \theta_j - \alpha \sum_{i=1}^m (f_{\theta}(x^i) - y^i) x_j^i$$

- Tipos de pre-procesado de datos:

$$X = X - \text{media}(X)$$

$$X = X - \text{devstd}(X)$$



Aumentado de datos (data augmentation)

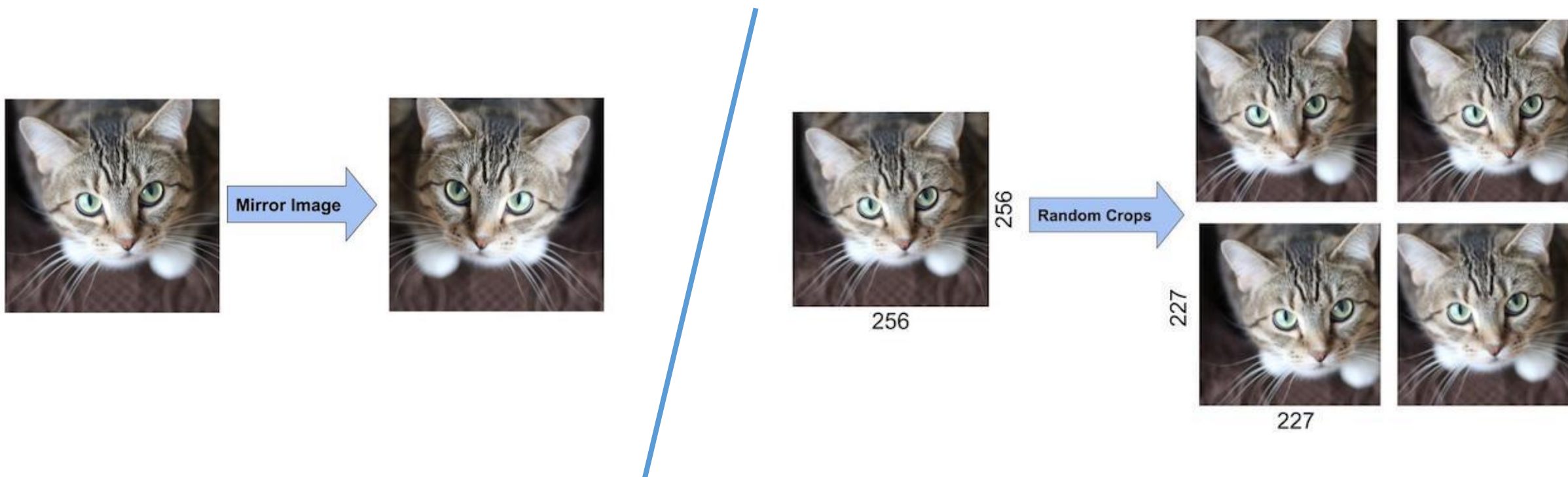
- El poder está en los **datos**
 - Cuanto más datos, mejor **generalización** en los modelos
 - La cantidad de datos es **limitada**: la recolección y etiquetado es un proceso costoso
- ¿Podemos dar a nuestra red neuronal más datos de los que tenemos?
 - ¡Sí! Aplicando **transformaciones** a los existentes.
 - Especialmente usado en **imágenes**.
 - Útil cuando tenemos pocos datos.
 - Aumenta robustez de la red



Transform image

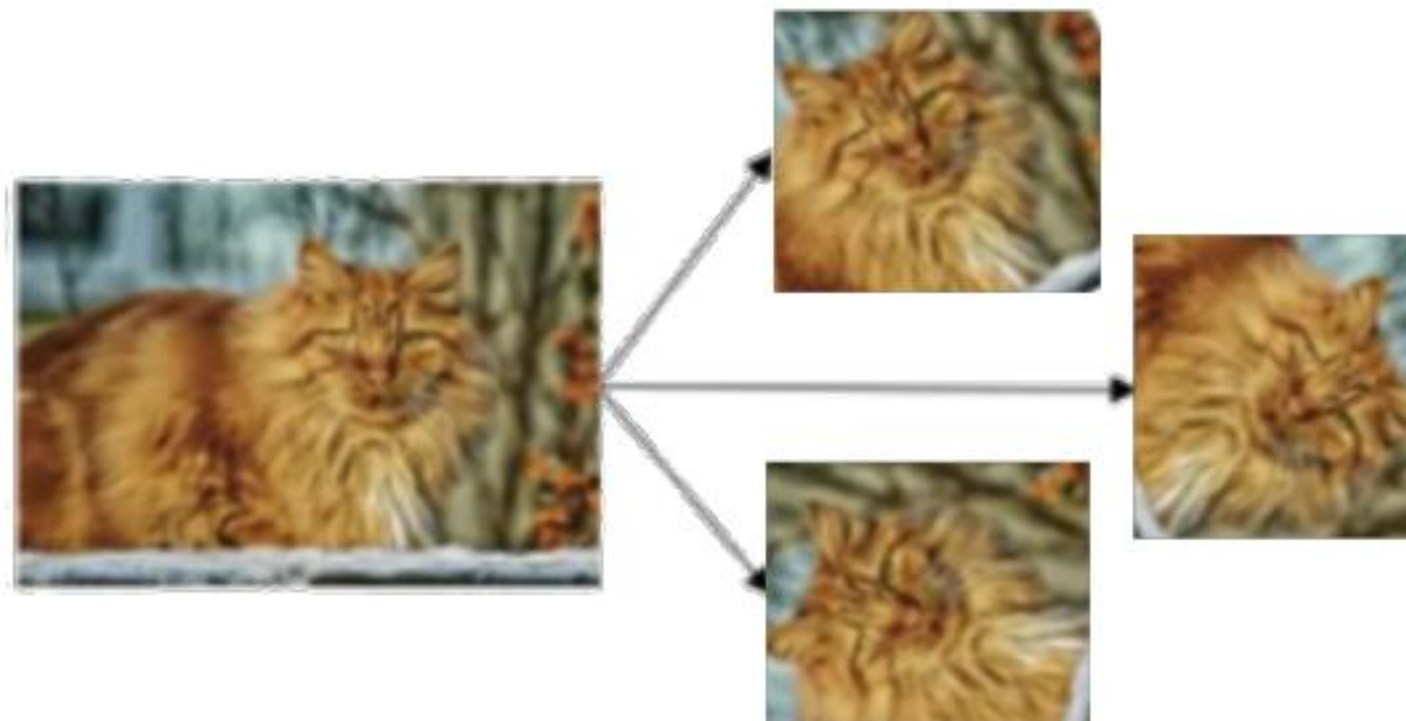
Aumentado de datos

- Transformaciones típicas: mirroring, cropping



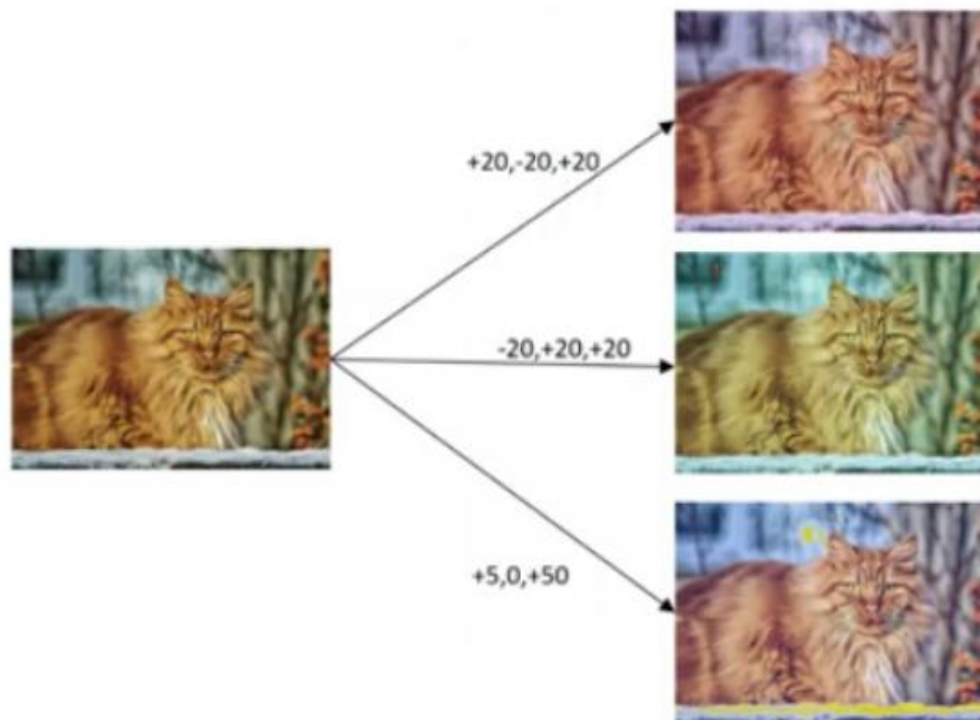
Aumentado de datos

- Transformaciones típicas: rotations



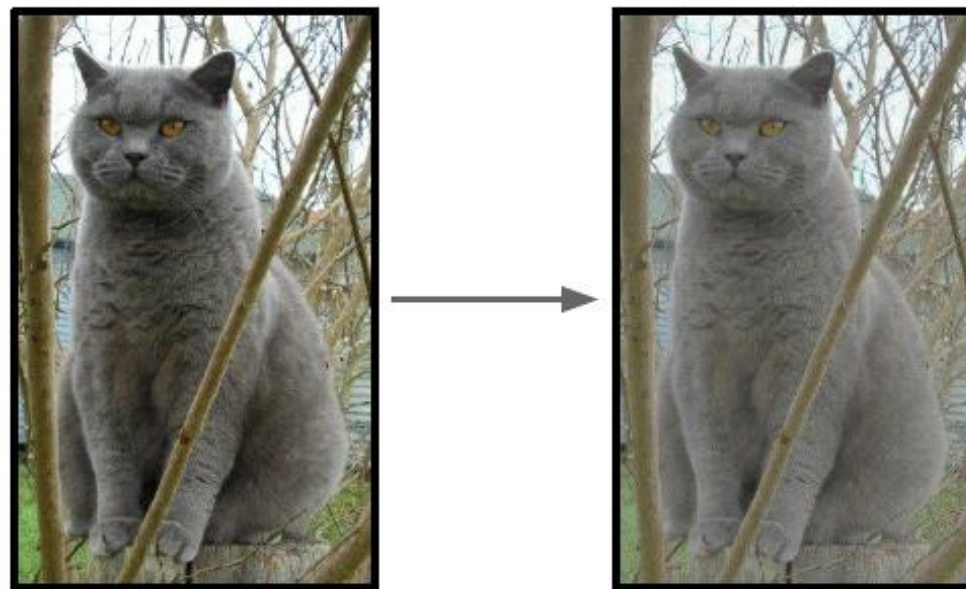
Aumentado de datos

- Transformaciones típicas: color shifting



Aumentado de datos

- Transformaciones típicas: color jittering
 - Introducido en el artículo de ResNet 2012
 - Algo complejo, basado en aplicar PCA

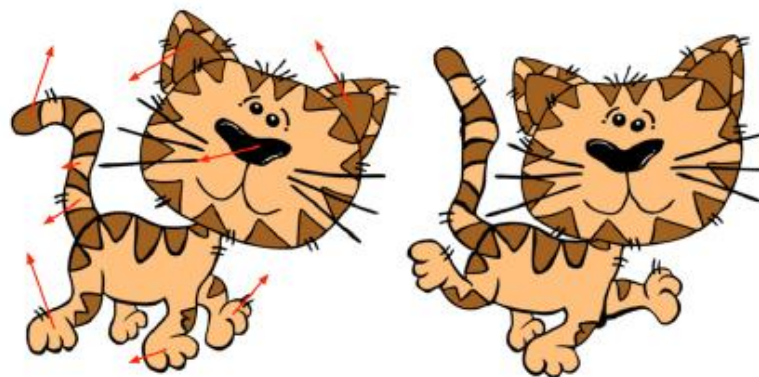


Aumentado de datos

- Transformaciones típicas:
 - Shearing

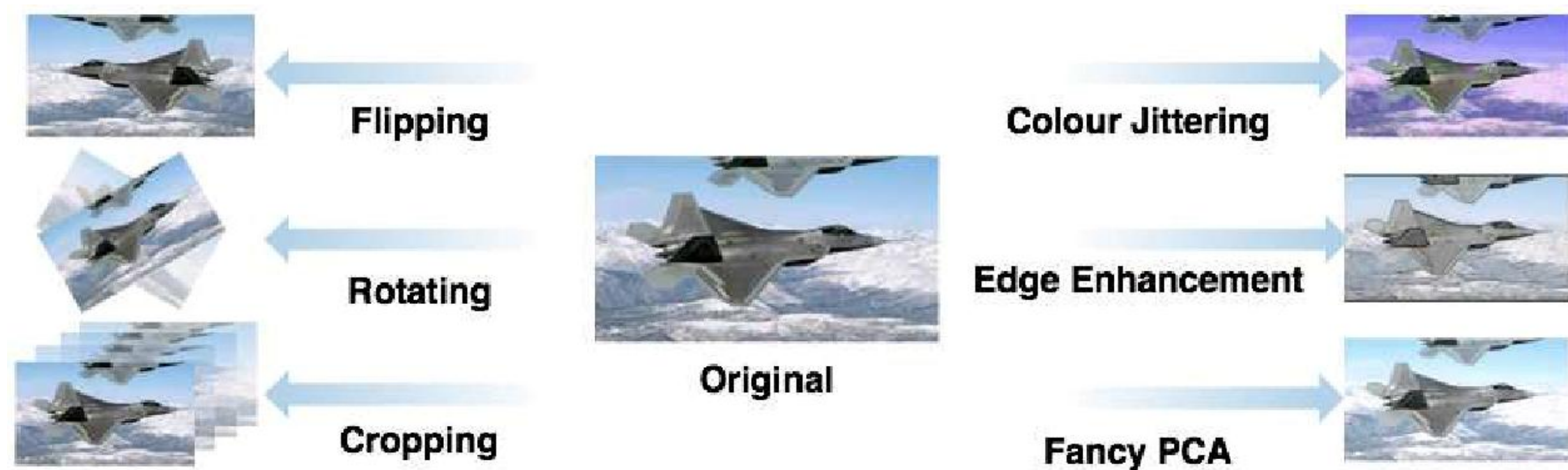


- Warping



Aumentado de datos

- Transformaciones típicas: muchas más, siendo creativo



Recapitulación

- Podemos también modificar los **datos** para evitar el **sobreajuste**
- Es importante primero **preprocesar** los datos:
 - Normalizarlos, centrarlos en cero y/o estandarizarlos
- Se puede **aumentar** la cantidad de datos aplicando transformaciones:
 - Imágenes, pocos datos.