

Fundamentos de Aprendizaje Automático y Reconocimiento de Patrones

Martín Rocamora, Guillermo Carbajal

Contiene material del curso *Aprendizaje Profundo para Visión Artificial*
(Delbracio, Lezama, Carbajal)

Clase 19

Tabla de contenido

Redes neuronales y aprendizaje profundo

- Aproximación versus generalización

- Regularización y validación

- Redes neuronales profundas

- Redes neuronales convolucionales

- Redes neuronales recurrentes

Tabla de contenido

Redes neuronales y aprendizaje profundo

Aproximación versus generalización

Regularización y validación

Redes neuronales profundas

Redes neuronales convolucionales

Redes neuronales recurrentes

Aproximación versus generalización

Teorema de aproximación universal

Una red neuronal prealimentada (feed-forward) con una única capa oculta y un número finito de neuronas, puede aproximar cualquier función continua en un espacio compacto de \mathbb{R}^n (Cybenko, 1989; Hornik 1991)

Aproximación versus generalización

Teorema de aproximación universal

Una red neuronal prealimentada (feed-forward) con una única capa oculta y un número finito de neuronas, puede aproximar cualquier función continua en un espacio compacto de \mathbb{R}^n (Cybenko, 1989; Hornik 1991)

- ▶ no establece como aprender los parámetros
- ▶ el ancho de la red debe ser exponencialmente grande
- ▶ aproximación universal para redes profundas de ancho acotado (*Lu et al., 2017*)

Aproximación versus generalización

Teorema de aproximación universal

Una red neuronal prealimentada (feed-forward) con una única capa oculta y un número finito de neuronas, puede aproximar cualquier función continua en un espacio compacto de \mathbb{R}^n (Cybenko, 1989; Hornik 1991)

- ▶ no establece como aprender los parámetros
- ▶ el ancho de la red debe ser exponencialmente grande
- ▶ aproximación universal para redes profundas de ancho acotado (*Lu et al., 2017*)

Generalización

Las redes neuronales tienen tendencia al **sobreajuste**.

Aproximación versus generalización

Teorema de aproximación universal

Una red neuronal prealimentada (feed-forward) con una única capa oculta y un número finito de neuronas, puede aproximar cualquier función continua en un espacio compacto de \mathbb{R}^n (Cybenko, 1989; Hornik 1991)

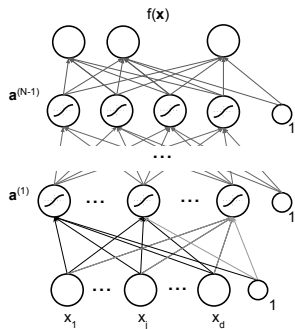
- ▶ no establece como aprender los parámetros
- ▶ el ancho de la red debe ser exponencialmente grande
- ▶ aproximación universal para redes profundas de ancho acotado (Lu et al., 2017)

Generalización

Las redes neuronales tienen tendencia al **sobreajuste**.

- ▶ resultados teóricos sobre error de generalización (d_{VC})
- ▶ **validación** y **regularización** para prevenir sobreajuste

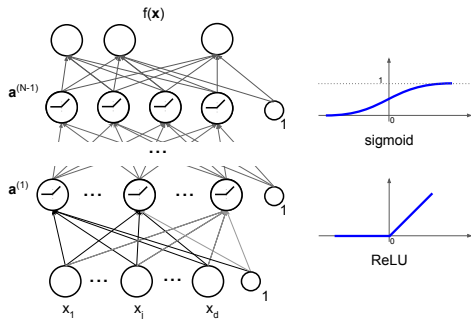
Redes neuronales profundas: múltiples capas



$$f(\mathbf{x}) = (g \circ a^{(L-1)} \circ \dots \circ a^{(1)})(\mathbf{x})$$

Redes neuronales profundas: funciones de activación

- ▶ el gradiente puede desvanecerse o explotar en el entrenamiento
- ▶ inicialización inadecuada y función de activación que satura



- ▶ la función ReLU: $h(x) = \max(0, x)$ da mejores resultados
- ▶ **Batch Normalization**: escalado y corrimiento de las entradas de cada capa (con parámetros aprendidos).

Tabla de contenido

Redes neuronales y aprendizaje profundo

Aproximación versus generalización

Regularización y validación

Redes neuronales profundas

Redes neuronales convolucionales

Redes neuronales recurrentes

Regularización y validación

- ▶ Weight decay:

Error con penalización sobre la complejidad del modelo.

$$E_{aug}(w) = E_{in}(w) + \frac{\lambda}{N} \sum_{i,j,l} (w_{ij}^{(l)})^2$$

Regularización y validación

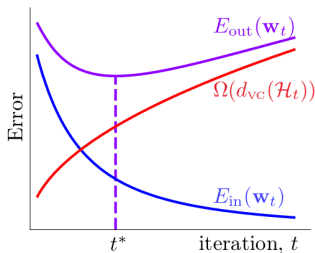
► Weight decay:

Error con penalización sobre la complejidad del modelo.

$$E_{aug}(\mathbf{w}) = E_{in}(\mathbf{w}) + \frac{\lambda}{N} \sum_{i,j,l} (w_{ij}^{(l)})^2$$

► Early stopping:

Se detiene el entrenamiento antes de sobreajustar al conjunto de entrenamiento, evaluando en el conjunto de validación.



Regularización y validación

▶ Dropout

Ignorar aleatoriamente algunas neuronas (10 a 50%) en cada paso de entrenamiento. Luego del entrenamiento se usan todas las neuronas. Se obtiene una red menos sensible a cambios en la entrada y menos sobreajustada.

Regularización y validación

► Dropout

Ignorar aleatoriamente algunas neuronas (10 a 50%) en cada paso de entrenamiento. Luego del entrenamiento se usan todas las neuronas. Se obtiene una red menos sensible a cambios en la entrada y menos sobreajustada.

► Data augmentation

Aumentar artificialmente los datos de entrenamiento mediante transformaciones (realistas) de los datos de entrada.

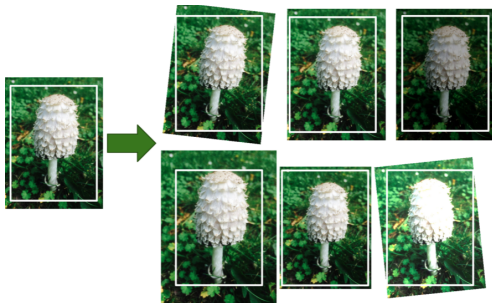


Tabla de contenido

Redes neuronales y aprendizaje profundo

Aproximación versus generalización

Regularización y validación

Redes neuronales profundas

Redes neuronales convolucionales

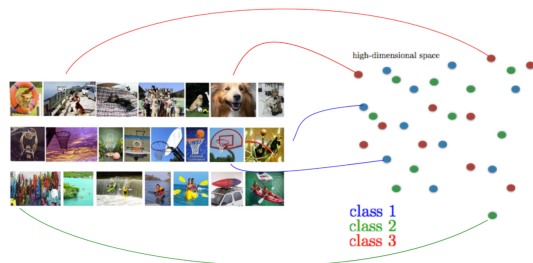
Redes neuronales recurrentes

Redes neuronales profundas: aprendizaje profundo

Una clase de representaciones paramétricas no lineales capaces de codificar características (o conocimiento) del problema y de ser optimizadas de forma eficiente (a enorme escala) usando métodos de descenso por gradiente estocástico.

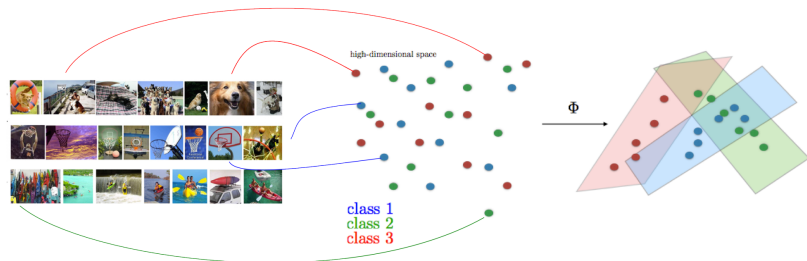
Redes neuronales profundas: aprendizaje profundo

Una clase de representaciones paramétricas no lineales capaces de codificar características (o conocimiento) del problema y de ser optimizadas de forma eficiente (a enorme escala) usando métodos de descenso por gradiente estocástico.



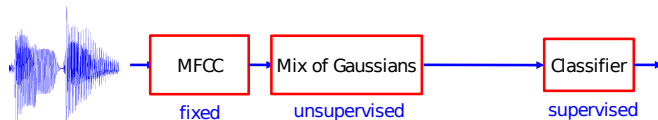
Redes neuronales profundas: aprendizaje profundo

Una clase de representaciones paramétricas no lineales capaces de codificar características (o conocimiento) del problema y de ser optimizadas de forma eficiente (a enorme escala) usando métodos de descenso por gradiente estocástico.

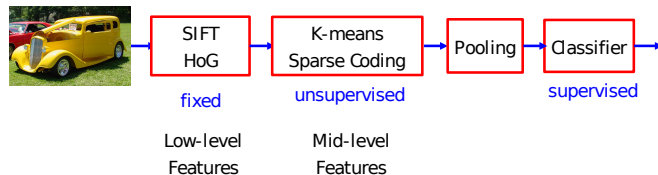


Arquitecturas de reconocimiento de patrones

Reconocimiento de audio, principios de los 90 hasta el 2011:

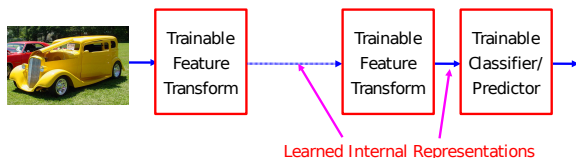


Reconocimiento de imágenes, principios del 2000 hasta el 2011:



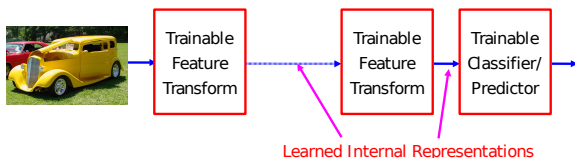
Ejemplo tomado de charlas de Y. Lecun.

Representaciones jerárquicas adaptivas



- ▶ Una jerarquía de representaciones (capas) con **mayores niveles de abstracción**
- ▶ Cada capa es una representación de características adaptiva (entrenable)

Representaciones jerárquicas adaptivas



- ▶ Una jerarquía de representaciones (capas) con **mayores niveles de abstracción**
- ▶ Cada capa es una representación de características adaptiva (entrenable)
- ▶ Imágenes:
 - ▶ píxeles → bordes → partes → objetos

Tabla de contenido

Redes neuronales y aprendizaje profundo

Aproximación versus generalización

Regularización y validación

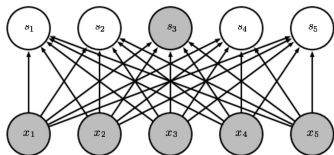
Redes neuronales profundas

Redes neuronales convolucionales

Redes neuronales recurrentes

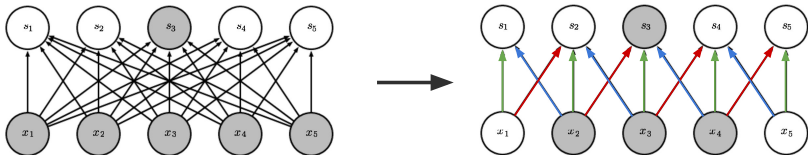
Redes neuronales convolucionales (CNN)

- ▶ Para datos en altas dimensiones, las redes neuronales *totalmente conectadas*, tal como vimos hasta ahora requieren un gran número de parámetros
- ▶ Por ejemplo, procesar imágenes de Imagenet (tamaño $224 \times 224 \times 3$) para clasificar entre 1000 clases, con una sola capa requeriría 150 millones de parámetros.

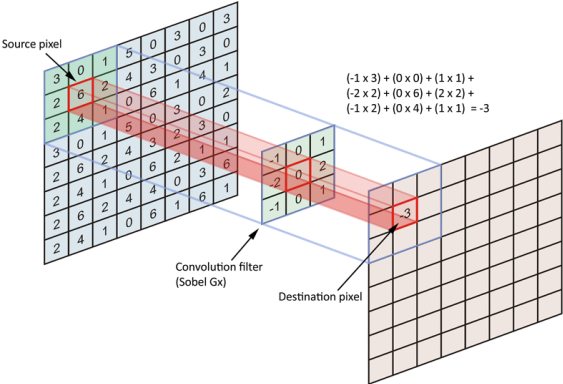


Redes neuronales convolucionales (CNN)

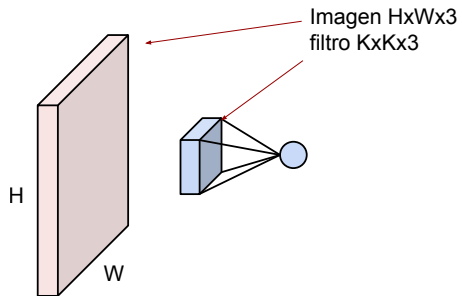
- ▶ Para datos en altas dimensiones, las redes neuronales *totalmente conectadas*, tal como vimos hasta ahora requieren un gran número de parámetros
- ▶ Por ejemplo, procesar imágenes de Imagenet (tamaño $224 \times 224 \times 3$) para clasificar entre 1000 clases, con una sola capa requeriría 150 millones de parámetros.
- ▶ Las redes de convolución surgen para evitar este problema haciendo cálculos más eficientes y explotando invarianzas inherentes a las imágenes



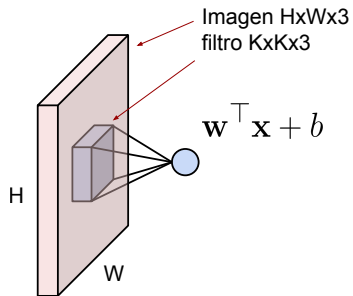
Convolución



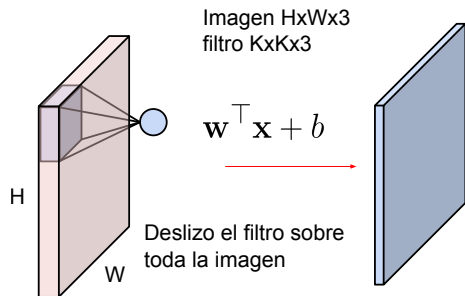
Capa de convolución



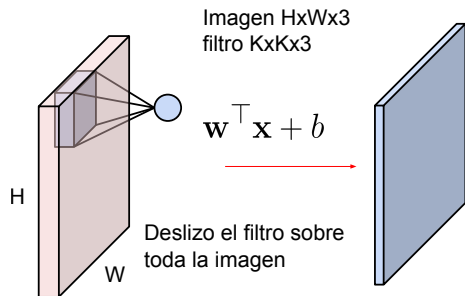
Capa de convolución



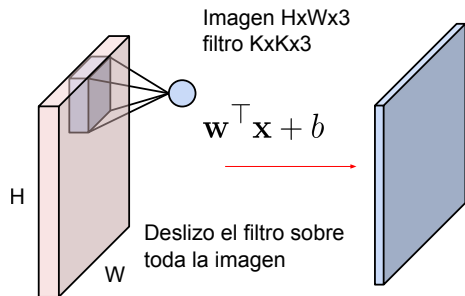
Capa de convolución



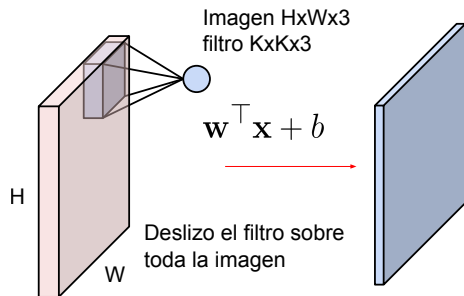
Capa de convolución



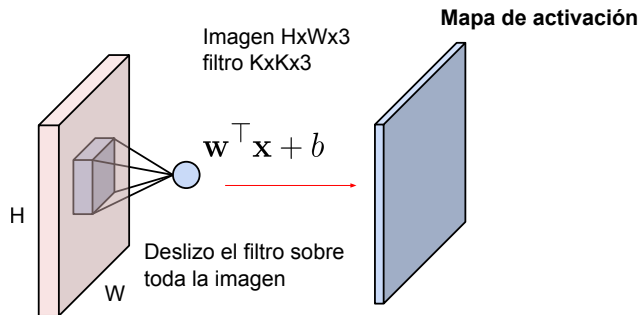
Capa de convolución



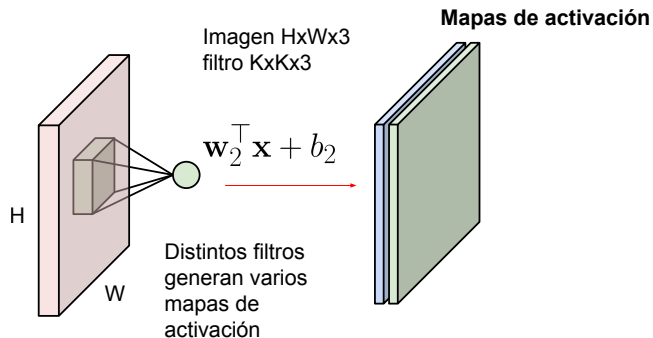
Capa de convolución



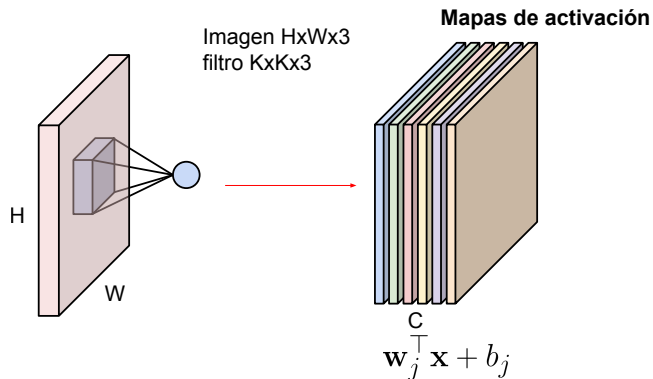
Capa de convolución



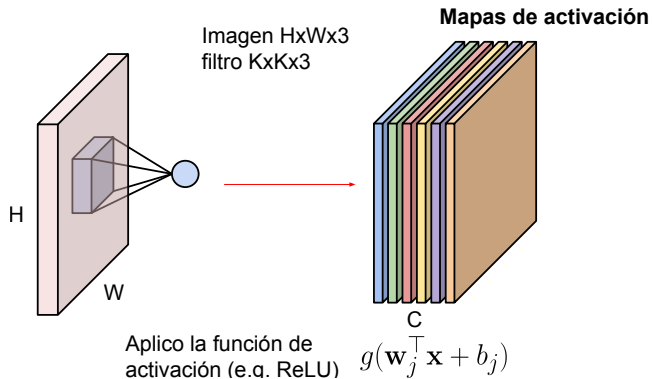
Capa de convolución



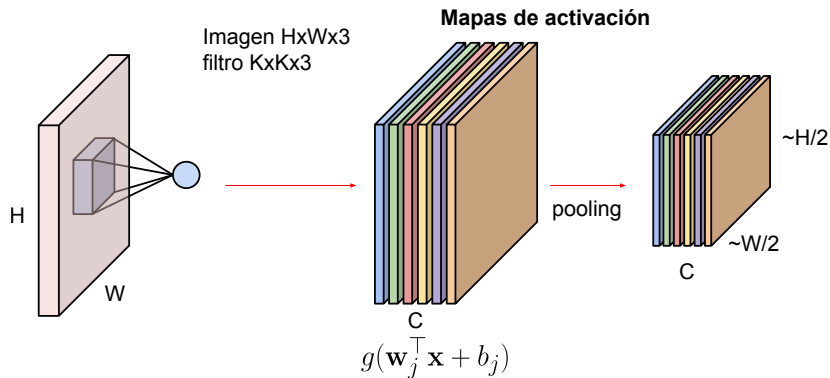
Capa de convolución



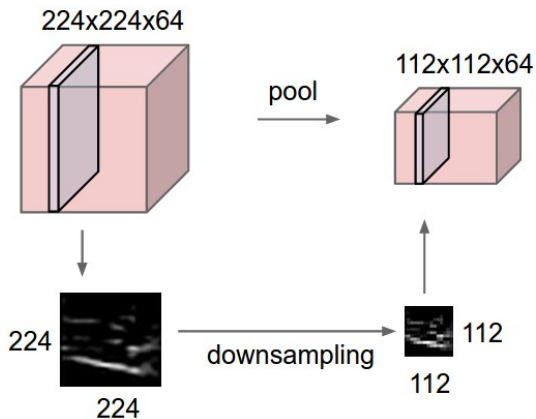
Capa de convolución



Capa de convolución



Pooling



Arquitectura típica de una CNN

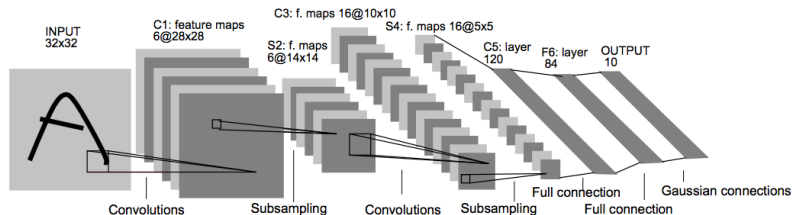
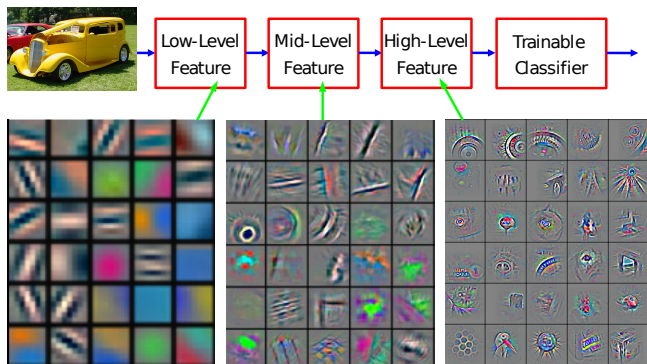


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

- ▶ Arquitectura LeNet-5, propuesta por *Yann Lecun et al. 1998*

¿Qué filtros aprende una CNN?



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

¿Qué filtros aprende una CNN?

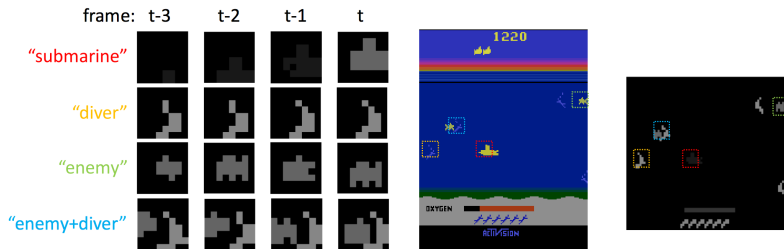


Figure 2: Visualization of the first-layer features learned from Seaquest. (Left) visualization of four first-layer filters; each filter covers four frames, showing the spatio-temporal template. (Middle) a captured screen. (Right) gray-scale version of the input screen which is fed into the CNN. Four filters were color-coded and visualized as dotted bounding boxes at the locations where they get activated. This figure is best viewed in color.

Tabla de contenido

Redes neuronales y aprendizaje profundo

Aproximación versus generalización

Regularización y validación

Redes neuronales profundas

Redes neuronales convolucionales

Redes neuronales recurrentes

Redes neuronales recurrentes (RNN)

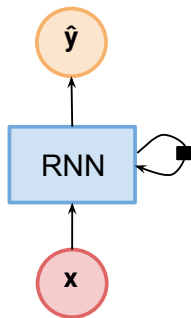
- ▶ para problemas secuenciales donde
 - ▶ largo de las secuencias de entrada/salida variable
 - ▶ quiero guardar información de la *historia*
- ▶ guardan un estado oculto h_t que depende de la entrada actual y el estado anterior h_{t-1}

$$h_t = f_w(h_{t-1}, x_t)$$

- ▶ ejemplo básico de RNN

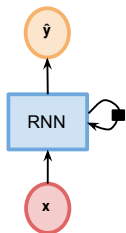
$$h_t = \tanh(W_{hh}h_{t-1} + W_{hx}x_t)$$

$$\hat{y}_t = W_{hy}h_t$$



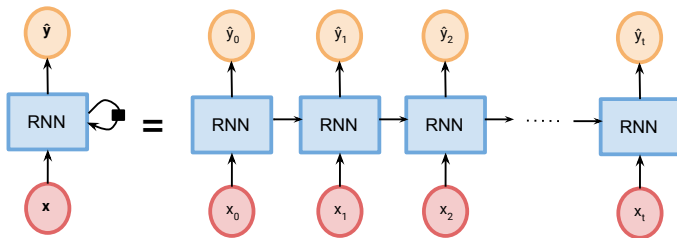
Redes neuronales recurrentes (RNN)

- ▶ Una red neuronal recurrente puede pensarse como múltiples copias de la misma red donde cada una le pasa un mensaje (estado) al sucesor



Redes neuronales recurrentes (RNN)

- ▶ Una red neuronal recurrente puede pensarse como múltiples copias de la misma red donde cada una le pasa un mensaje (estado) al sucesor

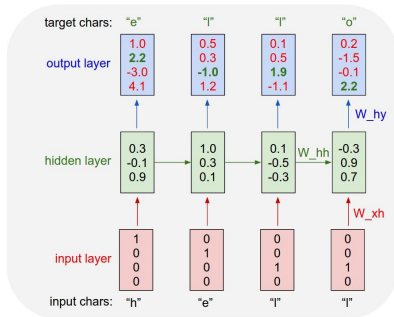


Ejemplo: modelado de lenguaje a nivel de carácter

Example: Character-level Language Model

Vocabulary:
[h,e,l,o]

Example training
sequence:
"hello"



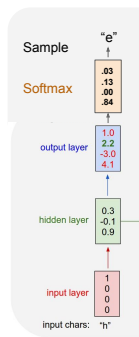
Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: modelado de lenguaje a nivel de carácter

Example: Character-level Language Model Sampling

Vocabulary:
[h,e,l,o]

At test-time sample
characters one at a time,
feed back to model

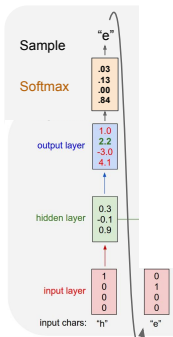


Ejemplo: modelado de lenguaje a nivel de carácter

Example: Character-level Language Model Sampling

Vocabulary:
[h,e,l,o]

At test-time sample
characters one at a time,
feed back to model

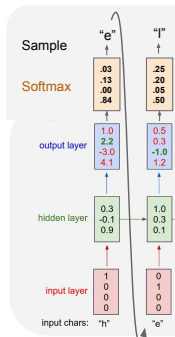


Ejemplo: modelado de lenguaje a nivel de carácter

Example: Character-level Language Model Sampling

Vocabulary:
[h,e,l,o]

At test-time sample
characters one at a time,
feed back to model

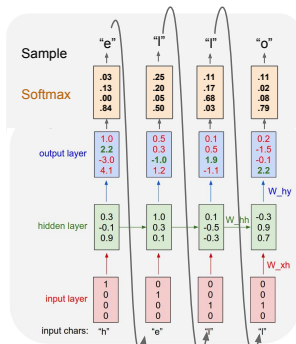


Ejemplo: modelado de lenguaje a nivel de carácter

Example: Character-level Language Model Sampling

Vocabulary:
[h,e,l,o]

At test-time sample
characters one at a time,
feed back to model



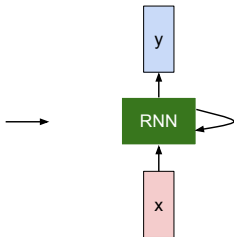
Ejemplo: modelado de lenguaje a nivel de carácter

THE SONNETS

by William Shakespeare

From faintest creatures we desire income,
That thereby beauty's rose might never die,
But as the ripen should by time decrease,
His tender heir might bear his memory;
But thou, contracted to thine own bright eyes,
Feed'st thy light's flame with self-substantial fuel,
Making a famine where abundance lies,
Thyself thy foe, to thy sweet self too cruel:
Thou that art now the world's fresh ornament,
And only herald to the gaudy spring,
Wilt thou time-consume that hast thee to decay,
And tender chaste make's waste in ragging;
Hap the world, or else the glutton be,
To eat the world's due, by the grave and thee.

When forty winters shall besiege thy brow,
And dig deep trenches in thy beauty's field,
Thy youth's proud livery so gazed on now,
Will be a tattered weed of small worth held;
Then being asked, where all this beauty lies,
Where all the treasure of thy lusty days,
To say, within whose deep sad eyes,
Were an all-eating shame, and thriftless grace.
How much more praise deserve'd thy beauty's use,
If thou couldst answer 'This fair child of mine
Shall sum my cost, and make my old excuse,'
Proving his beauty by succession thine!
This were to be new made when thou art old,
And see thy blood warm when thou feel'st it cold.



Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: modelado de lenguaje a nivel de carácter

at first:

tyntd-iafhatawiaoihrdemot lytdws e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrqd t o idoe ns,smtt h ne etie h,hregtrs nigtkie,aoaenns lng

↓
train more

"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."

↓
train more

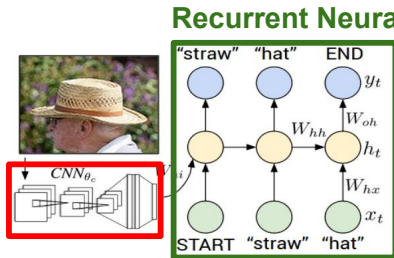
Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and offer.

↓
train more

"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.

Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes



Convolutional Neural Network

Andrej Karpathy Li Fei-Fei

"Deep Visual-Semantic Alignments for Generating Image Descriptions"

Ejemplo: generación automática de leyenda en imágenes

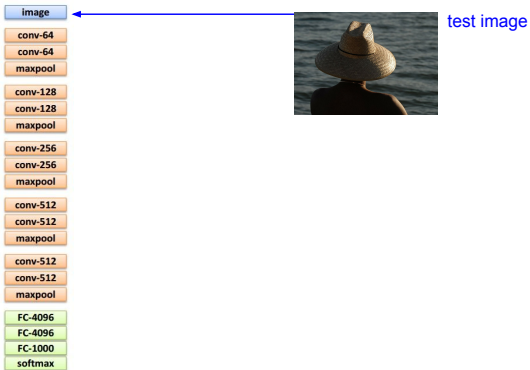


test image

This image is CC0 public domain

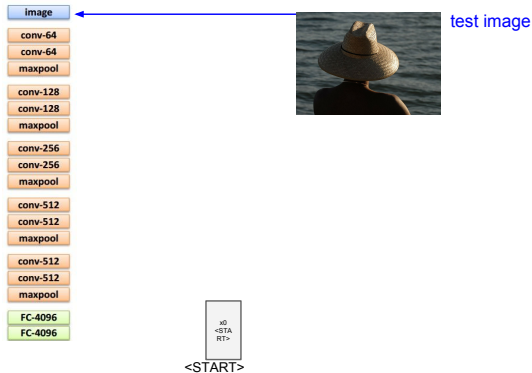
Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes



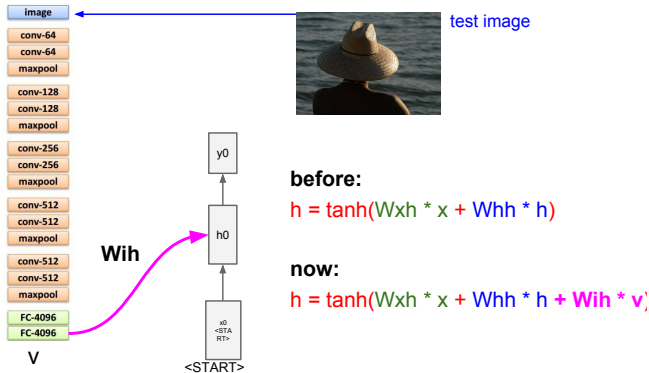
Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes



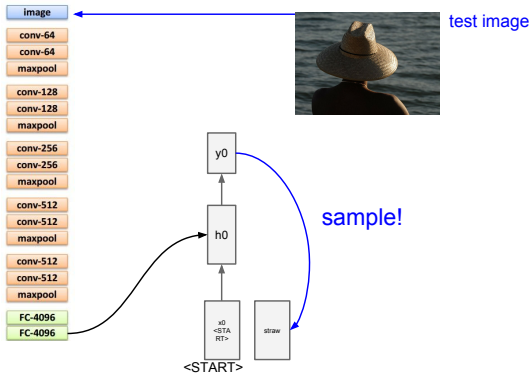
Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes



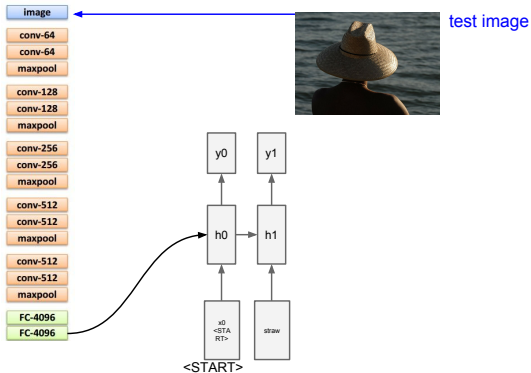
Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes



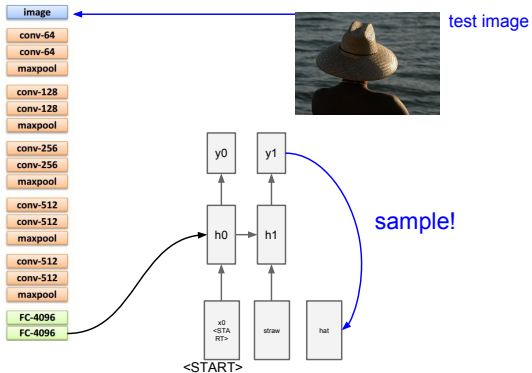
Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes



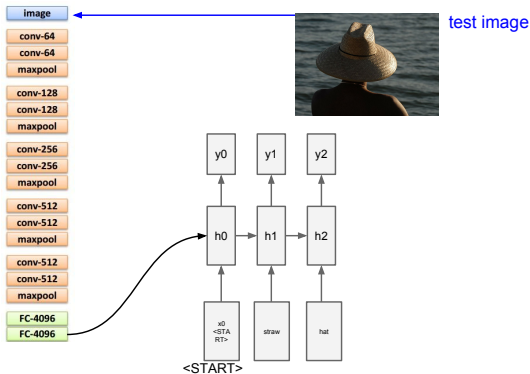
Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes



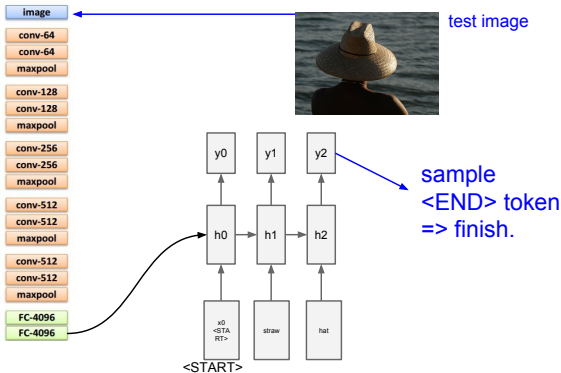
Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes



Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes



Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes

Image Captioning: Example Results

Captions generated using [pycocotools](#)
All images are [CC-BY Public Domain](#)
all authors are listed in the text
unless stated otherwise



A cat sitting on a suitcase on the floor



A cat is sitting on a tree branch



A dog is running in the grass with a frisbee



A white teddy bear sitting in the grass



Two people walking on the beach with surfboards



A tennis player in action on the court



Two giraffes standing in a grassy field



A man riding a dirt bike on a dirt track

Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung

Ejemplo: generación automática de leyenda en imágenes

Image Captioning: Failure Cases

Captions generated using [Caption2Vec](#)
All images are [CC0 Public Domain](#) for
most, but some are under various licenses



A woman is holding a cat in her hand



A person holding a computer mouse on a desk



A woman standing on a beach holding a surfboard



A bird is perched on a tree branch



A man in a baseball uniform throwing a ball

Slides tomadas de [cs231n](#) (Stanford) - Fei-Fei Li & Justin Johnson & Serena Yeung