Deep Learning

Episodio 4: La Operación de Convolución

Fernando Gama

Escuela de Graduados en Ingeniería Informática y Sistemas, Facultad de Ingeniería, UBA

28 de Julio de 2022

Escenas del capítulo anterior...

- Minimizar el riesgo empírico usando gradiente descendente
  $\Rightarrow$ Implica que la función es derivable $\Rightarrow$ Costo vs. Pérdida
- No necesariamente necesitamos encontrar el mínimo, sino bajar el costo

- Muestras de validación $\Rightarrow$ Sirven para observar el overfitting

- Regularización $\Rightarrow$ Técnicas para mejorar la generalización del modelo
  $\Rightarrow$ Penalidades $\Rightarrow$ Agregan un término a la función de pérdida
  $\Rightarrow$ Parada anticipada $\Rightarrow$ Fácil implementación, requiere muestras de validación
  $\Rightarrow$ Dropout $\Rightarrow$ Apagar neuronas de manera aleatoria $\Rightarrow$ Evitar memorización

Señales y Sistemas

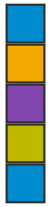
Convolución

Respuesta en Frecuencia

Convoluciones en Imágenes

Datos como Vectores

- Los datos los consideramos vectores $\mathbf{x} \in \mathbb{R}^N$
  - $\Rightarrow$ Asumimos que los elementos del vector no guardan relación entre sí

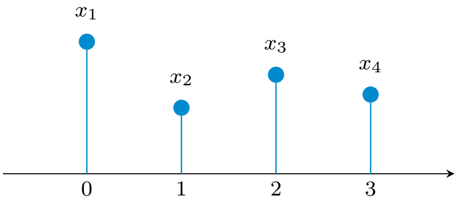


$\mathbf{x} \in \mathbb{R}^N$

- ¿Qué pasa si el vector representa algún otro tipo de dato que sí guarda relación entre los elementos?
  - $\Rightarrow$ ¿Qué hacemos?
  - $\Rightarrow$ ¿Podemos usar esa relación para mejorar nuestros algoritmos de aprendizaje?

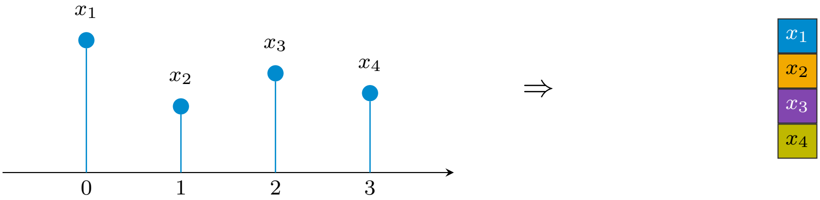
Señales

- Un tipo de datos que son muy comunes son los datos que dependen del tiempo
- Sea $\mathcal{N} = \{0, \ldots, N-1\}$ un conjunto de N índices temporales discretos
- Definimos una señal como $\mathrm{x} : \mathcal{N} \to \mathbb{R}$ $\Rightarrow$ A cada índice temporal le asignamos un escalar



Señales como Vectores

- Podemos describir la señal como un vector N-dimensional



- El primer elemento del vector corresponde al valor de la señal en tiempo 0
  $\Rightarrow$ El segundo elemento al tiempo 1, y así sucesivamente
- Nosotros sabemos que los elementos del vector están relacionados
  $\Rightarrow$ Pero esta relación no aparece naturalmente en el vector $\Rightarrow$ ¿Cómo lo recuperamos?

Sistemas

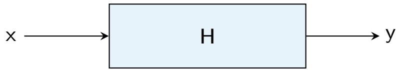
- Lo que queremos hacer es transformar los datos de alguna manera apropiada

  $\Rightarrow$ Si los datos tienen alguna estructura particular $\Rightarrow$ Explotar la estructura en la transformación

- Sistema $\Rightarrow$ Una función que toma una señal y devuelve una señal
- Si definimos $\mathbb{X}_N$ como el espacio de señales $\mathbb{X}_N = \{\mathsf{x} : \mathcal{N} \to \mathbb{R}\}$

  $\Rightarrow$ Un sistema es una función $\mathsf{H} : \mathbb{X}_N \to \mathbb{X}_N$
- El procesamiento de la señal a través del sistema H tiene que explotar la estructura de los datos

  $\Rightarrow$ A través del uso de sistemas pertinentes recuperamos la noción de tiempo



Ejemplos de Sistemas

fgama@fi.uba.ar

- Sistema lineal: $\mathsf{y}(n) = \mathsf{H}(\mathsf{x}) = a\mathsf{x}(n)$ para algún $a \in \mathbb{R}$
  - $\Rightarrow$ Como vectores: $\mathbf{x} = [x_1, x_2, x_3, x_4] \mapsto \mathbf{y} = [ax_1, ax_2, ax_3, ax_4]$
- Sistema no-lineal: $\mathsf{y}(n) = \mathsf{H}(\mathsf{x}) = (\mathsf{x}(n))^2$ $\Rightarrow$ Aplicamos la función a cada instante de tiempo
  - $\Rightarrow$ Como vectores: $\mathbf{x} = [x_1, x_2, x_3, x_4] \mapsto \mathbf{y} = [x_1^2, x_2^2, x_3^2, x_4^2]$
- Estos son ejemplos de sistemas sin memoria $\Rightarrow$ La noción de tiempo no aparece de manera explícita

- Sistema con demora unitaria: $\mathsf{y}(n) = \mathsf{H}(\mathsf{x}) = \mathsf{x}(n-1)$
  - $\Rightarrow$ Como vectores: $\mathbf{x} = [x_1, x_2, x_3, x_4] \mapsto [?, x_1, x_2, x_3]$
- Promediador móvil: $\mathsf{y}(n) = \mathsf{H}(\mathsf{x}) = \mathsf{x}(n) + \mathsf{x}(n-1)$
  - $\Rightarrow$ Como vectores: $\mathbf{x} = [x_1, x_2, x_3, x_4] \mapsto [x_1+?, x_2 + x_1, x_3 + x_2, x_4 + x_3]$
- ¿Qué se hace en estos casos donde necesito datos temporales que no tengo?
  - $\Rightarrow$ Hay muchas alternativas: asumir que son cero (*zero-padding*), reducir el tamaño de salida, ...

Señales y Sistemas

Convolución

Respuesta en Frecuencia

Convoluciones en Imágenes

- Sistema Lineal $\Rightarrow$ Combinación lineal a la entrada produce combinación lineal a la salida

$$\mathsf{H}\big(a_1\mathsf{x}_1 + a_2\mathsf{x}_2 + \cdots + a_K\mathsf{x}_K\big) = a_1\mathsf{H}(\mathsf{x}_1) + a_2\mathsf{H}(\mathsf{x}_2) + \cdots + a_K\mathsf{H}(\mathsf{x}_K)$$

- Operación de demora o desplazamiento $\mathsf{y}(n) = \mathsf{D}_k(\mathsf{x}) = \mathsf{y}(n-k)$
- Sistema invariante en el tiempo

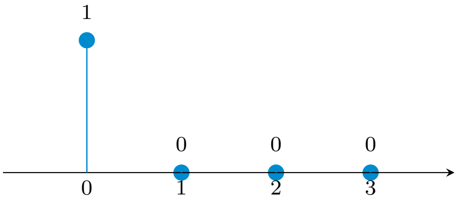
$$\mathsf{H}\big(\mathsf{D}_k(\mathsf{x})\big) = \mathsf{D}_k\big(\mathsf{H}(\mathsf{x})\big)$$

$\Rightarrow$ Desplazar la señal de entrada genera un desplazamiento igual en la salida

Si $\mathsf{y}(n) = \mathsf{H}(\mathsf{x})$ es la salida, entonces $\mathsf{H}\big(\mathsf{x}(n-k)\big) = \mathsf{H}\big(\mathsf{D}_k(\mathsf{x})\big) = \mathsf{D}_k\big(\mathsf{H}(\mathsf{x})\big) = \mathsf{D}_k\big(\mathsf{y}\big) = \mathsf{y}(n-k)$

- Definimos la señal conocida como *delta* (también se conoce como impulso)

$$\delta(n) = \begin{cases} 1 & \text{si } n = 0 \\ 0 & \text{en otro caso} \end{cases}$$

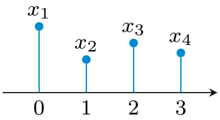


- Como vector, se puede escribir como $\boldsymbol{\delta} = [1, 0, \ldots, 0] \in \mathbb{R}^N$

Descomposición de una Señal

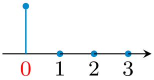
- Explotando las nociones de linealidad y desplazamiento, podemos descomponer cualquier señal

$$\mathsf{x} = \sum_{n=0}^{N-1} x_n \mathsf{D}_n(\delta) = x_0 D_0(\delta) + x_1 \mathsf{D}_1(\delta) + \cdots + x_{N-1} \mathsf{D}_{n-1}(\delta)$$

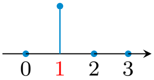


$=$

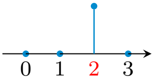
$x_1 \cdot$



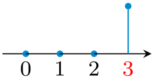
$+ \; x_2 \cdot$



$+ \; x_3 \cdot$

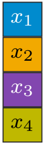


$+ \; x_4 \cdot$



Descomposición de una Señal

- Descomponer una señal como una combinación lineal de deltas desplazadas es obvio en vectores



$$= \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Filtrado con un Sistema LTI

- Usamos la descomposición para calcular la salida de un sistema LTI (lineal, invariante en el tiempo)

$$\mathsf{H}(\mathsf{x}) = \mathsf{H}\left(\sum_{n=0}^{N-1} x_n \mathsf{D}_n(\delta)\right) = \sum_{n=0}^{N-1} x_n \mathsf{D}_n\big(\mathsf{H}(\delta)\big)$$

- $\mathsf{H}(\delta)$ es aplicar el sistema LTI H a la señal δ

  $\Rightarrow$ Se conoce como respuesta impulsiva del filtro $\Rightarrow$ También es una señal $\mathsf{h} = \mathsf{H}(\delta)$

$$\mathsf{H}(\mathsf{x}) = \sum_{n=0}^{N-1} x_n \mathsf{D}_n\big(\mathsf{h}\big)$$

- La salida del filtro

  $\Rightarrow$ Se calcula como una combinación lineal de versiones desplazadas de la respuesta impulsiva

  $\Rightarrow$ Los coeficientes de la combinación son los valores de la señal de entrada

Filtrado con un Sistema LTI

- La respuesta impulsiva del filtro también es una señal h

$$\mathsf{H}(\mathsf{x}) = \sum_{n=0}^{N-1} x_n \mathsf{D}_n(\mathsf{h})$$

⇒ Esta señal caracteriza por completo al filtro

⇒ Conociendo h podemos calcular la salida para cualquier señal de entrada

- Podemos pensar la salida del filtro como poner h a la entrada de un filtro X con respuesta impulsiva x

$$\mathsf{X}(\mathsf{h}) = \sum_{n=0}^{N-1} h_n \mathsf{D}_n(\mathsf{x}) = \sum_{n=0}^{N-1} x_n \mathsf{D}_n(\mathsf{h}) = \mathsf{H}(\mathsf{x})$$

⇒ La operación para calcular la salida del filtro es conmutativa

⇒ La respuesta impulsiva h puede tener $K \ll N$ coeficientes ⇒ Se puede diseñar (o aprender)

‘

- La operación para calcular la salida del filtro se conoce como convolución

$$\mathsf{h} * \mathsf{x} = \sum_{k=0}^{K-1} h_k \mathsf{D}_k(\mathsf{x})$$

donde $\mathsf{h} = \sum_{k=0}^{K-1} h_k \mathsf{D}_k(\delta)$ es la respuesta impulsiva del filtro

- El resultado de la operación de convolución $\mathbb{X}_N \times \mathbb{X}_K \to \mathbb{X}_M$ es otra señal $\Rightarrow \mathsf{y} = \mathsf{h} * \mathsf{x}$

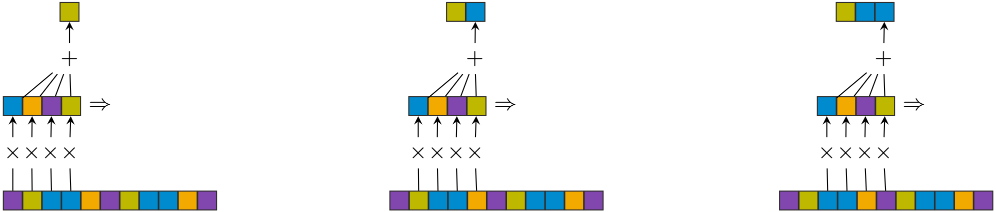
$\Rightarrow$ El elemento n de la señal de salida se puede calcular explícitamente como

$$\mathsf{y}(n) = \big(\mathsf{h} * \mathsf{x}\big)(n) = \sum_{k=0}^{K-1} \mathsf{h}(k)\, \mathsf{x}(n - k)$$

- Esta es la definición tradicional de convolución para señales en 1 dimensión
- Para otras señales el operador desplazamiento D_n puede tener otras formas

- La operación de convolución es: combinación lineal, desplazar, combinación lineal, desplazar, ...
- Tenemos dos señales: una es el filtro h y la otra son los datos x

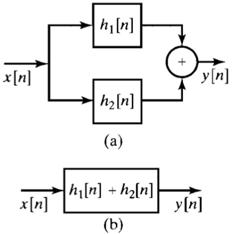
- La convolución implica una combinación lineal y un desplazamiento



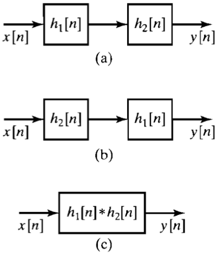
Propiedades de la Convolución

- La convolución es conmutativa $\Rightarrow \mathsf{h} * \mathsf{x} = \mathsf{x} * \mathsf{h}$ $\Rightarrow$ La noción de filtro y señal es intercambiable

  $\Rightarrow$ Conceptualmente, filtro $\neq$ señal (y porque hay otros dominios donde no conmutan)
- La convolución es lineal $\Rightarrow \mathsf{h} * (a_1\mathsf{x}_1 + a_2\mathsf{x}_2) = a_1(\mathsf{h} * \mathsf{x}_1) + a_2(\mathsf{h} * \mathsf{x}_2)$
- La convolución es asociativa $\Rightarrow \mathsf{h}_1 * (\mathsf{h}_2 * \mathsf{x}) = (\mathsf{h}_1 * \mathsf{h}_2) * \mathsf{x}$



Sistemas en paralelo



Sistemas en cascada

Dimensionalidad de la Convolución

- Tenemos un filtro h con K coeficientes, tenemos una señal x con N valores

  $\Rightarrow$ La convolución está bien definida en $N - K + 1$ valores

- No se rellena con ceros $\Rightarrow$ Tener cuidado de señales cada vez más pequeñas
- Se rellena con ceros de tal forma que la salida tenga tamaño N, igual que la entrada

  $\Rightarrow$ Los valores en los bordes tienen menos influencia en la salida $\Rightarrow$ Efectos de borde

Señales y Sistemas

Convolución

Respuesta en Frecuencia

Convoluciones en Imágenes

Autofunciones de los Sistemas LTI

- Consideremos la señal exponencial compleja $\mathsf{x}_t(n) = e^{\mathrm{j}(2\pi t/N)n}$ con $t \in \{0, 1, \ldots, N-1\}$
- Consideremos un sistema LTI con respuesta impulsiva $\mathbf{h} = [h_0, \ldots, h_{K-1}] \in \mathbb{R}^K$
- La salida se calcula mediante la operación de convolución

$$\mathsf{H}(\mathsf{x}(n)) = (\mathsf{h} * \mathsf{x}_t)(n) = \sum_{k=0}^{K-1} h_k e^{\mathrm{j}\frac{2\pi t}{N}(n-k)} = \left(\sum_{k=0}^{K-1} h_k e^{-\mathrm{j}\frac{2\pi t}{N}k} \right) e^{\mathrm{j}\frac{2\pi t}{N}n} = \tilde{h}_t e^{\mathrm{j}\frac{2\pi t}{N}n} = \tilde{h}_t \mathsf{x}_t(n)$$

$\Rightarrow$ La salida del sistema es igual a la entrada, pero reescalada $\Rightarrow \mathsf{H}(\mathsf{x}_t(n)) = \tilde{h}(t)\ \mathsf{x}_t(n)$

$\Rightarrow$ Resulta que la exponencial compleja $\mathsf{x}(n) = e^{\mathrm{j}(2\pi t/N)n}$ es una autofunción del sistema LTI

- El autovalor $\tilde{h}_t$ se conoce como la respuesta en frecuencia del filtro en la frecuencia t

$$\tilde{h}_t = \sum_{k=0}^{K-1} h_k e^{-\mathrm{j}\frac{2\pi t}{N}k}$$

- La respuesta en frecuencia del filtro a frecuencia t viene dada por

$$\tilde{h}_t = \sum_{k=0}^{N-1} h_k e^{-\mathrm{j}\,\frac{2\pi t}{N}\,k}$$

donde h_k son los valores de la respuesta impulsiva (que es una señal) y donde $h_k = 0$ para $k > K$

- Esta operación es un producto interno entre vectores $\Rightarrow \tilde{h}_t = \langle \mathbf{e}_t, \mathbf{h} \rangle = \mathbf{e}_t^{\mathsf{H}} \mathbf{h} \in \mathbb{C}$ tal que

$$\tilde{h}_t = \mathbf{e}_t^{\mathsf{H}} \mathbf{h} = \begin{bmatrix} e^{-\mathrm{j}\,\frac{2\pi t}{N}\,0} & e^{-\mathrm{j}\,\frac{2\pi t}{N}\,1} & \ldots & e^{-\mathrm{j}\,\frac{2\pi t}{N}\,(N-1)} \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{N-1} \end{bmatrix} = \sum_{k=0}^{N-1} h_k e^{-\mathrm{j}\,\frac{2\pi t}{N}\,k}$$

- Dijimos que había $t \in \{0, 1, \ldots, N-1\}$ valores distintos de frecuencia

  $\Rightarrow$ Cada uno se calcula como el producto interno $\tilde{h}_t = \mathbf{e}_t^{\mathsf{H}}\mathbf{h}$ $\Rightarrow$ El vector $\mathbf{h}$ no cambia

  $\Rightarrow$ Se puede escribir como un producto de una matriz con un vector $\Rightarrow$ Ponemos $\mathbf{e}_t^{\mathsf{H}}$ en las filas

$$\begin{bmatrix} \tilde{h}_0 \\ \tilde{h}_1 \\ \vdots \\ \tilde{h}_{N-1} \end{bmatrix} = \begin{bmatrix} e^{-\mathrm{j}\frac{2\pi 0}{N}0} & e^{-\mathrm{j}\frac{2\pi 0}{N}1} & \cdots & e^{-\mathrm{j}\frac{2\pi 0}{N}(N-1)} \\ e^{-\mathrm{j}\frac{2\pi 1}{N}0} & e^{-\mathrm{j}\frac{2\pi 1}{N}1} & \cdots & e^{-\mathrm{j}\frac{2\pi 1}{N}(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\mathrm{j}\frac{2\pi (N-1)}{N}0} & e^{-\mathrm{j}\frac{2\pi (N-1)}{N}1} & \cdots & e^{-\mathrm{j}\frac{2\pi (N-1)}{N}(N-1)} \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{N-1} \end{bmatrix}$$

- Obtenemos la respuesta en frecuencia del filtro para todas las frecuencias

$$\tilde{\mathbf{h}} = \mathbf{E}^{\mathsf{H}}\mathbf{h}$$

  $\Rightarrow$ Es una operación lineal, un producto de una matriz con un vector

Otra Descomposición de la Señal

- La respuesta en frecuencia se calcula como una transformación lineal de la señal

$$\tilde{\mathbf{h}} = \mathbf{E}^{\mathsf{H}}\mathbf{h}$$

- La matriz $\mathbf{E}^{\mathsf{H}}$ es una matriz unitaria (ortogonal) $\Rightarrow$ Su inversa es $(\mathbf{E}^{\mathsf{H}})^{-1} = \mathbf{E}$

  $\Rightarrow$ Podemos escribir la señal como una transformación lineal de la respuesta en frecuencia

$$\mathbf{E}\tilde{\mathbf{h}} = \mathbf{h}$$

- La señal a cada instante $\mathsf{h}(n)$ se puede escribir en una descomposición de exponenciales complejas

$$\mathsf{h}(n) = \sum_{t=0}^{N-1} \tilde{h}_t e^{\mathrm{j}\frac{2\pi t}{N}n}$$

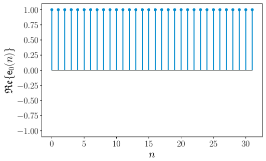
  $\Rightarrow$ Los elementos de la combinación lineal son la respuesta en frecuencia

Interpretaciones de Frecuencia

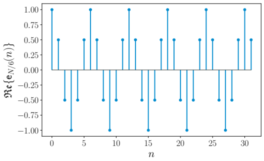
- Una exponencial compleja representa oscilaciones a frecuencia $2\pi t/N$

$$e^{\mathrm{j}\,\frac{2\pi t}{N}n} = \cos\left(\frac{2\pi t}{N}n\right) + \mathrm{j}\sin\left(\frac{2\pi t}{N}n\right)$$

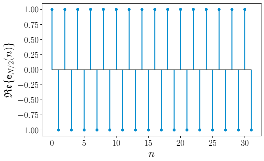
- Para $t = 0$ tenemos una señal constante $\Rightarrow \cos((2\pi 0/N)n) = 1$ y $\sin((2\pi 0/N)n) = 0$ para todo n
- A medida que t aumenta, las oscilaciones son cada vez más rápidas
- Las oscilaciones más rápidas de todas se dan para $t = N/2$
- Luego de $t = N/2$, las oscilaciones vuelven a bajar en frecuencia



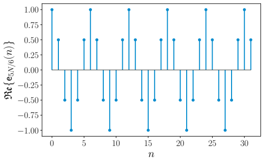
$t = 0$



$t = N/6$



$t = N/2$



$t = 5N/6$

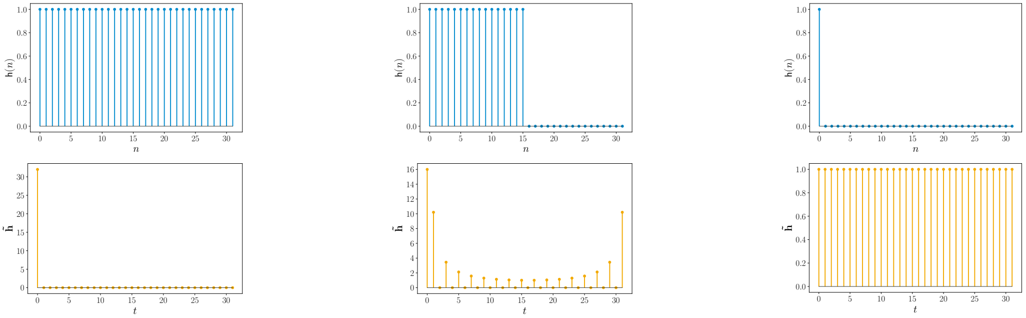
Interpretaciones de Frecuencia

- Cada componente de la respuesta en frecuencia se calcula como el producto interno

$$\tilde{h}_t = \langle \mathbf{e}_t, \mathbf{h} \rangle = \mathbf{e}_t^{\mathsf{H}} \mathbf{h}$$

$\Rightarrow$ El producto interno determina qué tanto se parecen los vectores

$\Rightarrow$ El producto $\mathbf{e}_t^{\mathsf{H}} \mathbf{h}$ me dice qué porción de la señal oscila a frecuencia t



El Efecto de la Convolución en Frecuencia

- Podemos escribir una señal como una descomposición de exponenciales complejas

$$\mathsf{h}(n) = \sum_{t=0}^{N-1} \tilde{h}_t e^{\mathrm{j}\,\frac{2\pi t}{N}\, n} \quad , \quad \mathsf{x}(n) = \sum_{t=0}^{N-1} \tilde{x}_t e^{\mathrm{j}\,\frac{2\pi t}{N}\, n}$$

- ¿Qué pasa con la respuesta en frecuencia de la convolución?

$$\mathsf{y}(n) = (\mathsf{h} * \mathsf{x})(n) = \sum_{t=0}^{N-1} \tilde{y}_t e^{\mathrm{j}\,\frac{2\pi t}{N}\, n}$$

- ¿Cuál es la relación entre las respuestas en frecuencia $\tilde{y}_t$, $\tilde{h}_t$ y $\tilde{x}_t$?

- La convolución se calcula como $\mathsf{y}(n) = \sum_{k=0}^{N-1} \mathsf{h}(k)\mathsf{x}(n-k)$
- Reemplazamos $\mathsf{h}(n)$ y $\mathsf{x}(n)$ por sus descomposiciones en exponenciales complejas

$$\mathsf{y}(n) = \sum_{k=0}^{N-1} \mathsf{h}(k)\mathsf{x}(n-k) = \sum_{k=0}^{N-1} \left(\sum_{\tau=0}^{N-1} \tilde{h}_\tau e^{\mathrm{j}\frac{2\pi\tau}{N}k} \right) \left(\sum_{t=0}^{N-1} \tilde{x}_t e^{\mathrm{j}\frac{2\pi t}{N}(n-k)} \right)$$

- Distribuimos $\sum_{t=0}^{N-1} \sum_{\tau=0}^{N-1} \sum_{k=0}^{N-1} \tilde{h}_\tau \tilde{x}_t e^{\mathrm{j}\frac{2\pi(\tau-t)}{N}k} e^{\mathrm{j}\frac{2\pi t}{N}n} = \sum_{t=0}^{N-1} \tilde{x}_t e^{\mathrm{j}\frac{2\pi t}{N}n} \sum_{\tau=0}^{N-1} \tilde{h}_\tau \sum_{k=0}^{N-1} e^{\mathrm{j}\frac{2\pi(\tau-t)}{N}k}$
- Usamos la propiedad de que las exponenciales complejas se cancelan a sí mismas

$$\sum_{\tau=0}^{N-1} \tilde{h}_\tau \sum_{k=0}^{N-1} e^{\mathrm{j}\frac{2\pi(\tau-t)}{N}k} = \begin{cases} N\tilde{h}_t & \text{si } \tau = t \\ 0 & \text{en otro caso} \end{cases}$$

$\Rightarrow$ Usando la suma geométrica: $\sum_{k=0}^{N-1} e^{-\mathrm{j}\frac{2\pi(\tau-t)}{N}k} = \begin{cases} \dfrac{1-e^{\mathrm{j}2\pi(t-\tau)}}{1-e^{\mathrm{j}\frac{2\pi(t-\tau)}{N}}} = 0 & \text{si } t \neq \tau \\ \sum_{k=0}^{N-1} 1 = N & \text{si } t = \tau \end{cases}$

El Efecto de la Convolución en Frecuencia

- Entonces, sabiendo que $\sum_{\tau=0}^{N-1} \tilde{h}_\tau \sum_{k=0}^{N-1} e^{\mathrm{j}\,\frac{2\pi(\tau-t)}{N}k} = \begin{cases} N\tilde{h}_t & \text{si } \tau = t \\ 0 & \text{en otro caso} \end{cases}$
- Lo usamos en la expresión de la convolución

$$\sum_{t=0}^{N-1} \tilde{x}_t e^{\mathrm{j}\,\frac{2\pi t}{N}n} \sum_{\tau=0}^{N-1} \tilde{h}_\tau \sum_{k=0}^{N-1} e^{\mathrm{j}\,\frac{2\pi(\tau-t)}{N}k} = N \sum_{t=0}^{N-1} \tilde{x}_t \tilde{h}_t e^{\mathrm{j}\,\frac{2\pi t}{N}n}$$

- Observamos que esto es una combinación lineal de exponenciales complejas
  - ⇒ Es la respuesta en frecuencia de $\mathsf{y}(n)$
  - ⇒ La respuesta en frecuencia a frecuencia t viene dada por

$$\tilde{y}_t = \tilde{h}_t \tilde{x}_t$$

El Efecto de la Convolución en Frecuencia

- La respuesta en frecuencia de una convolución es la multiplicación de las respuestas en frecuencia

Convolución en el tiempo $\Leftrightarrow$ Multiplicación en frecuencia

- Una vez que tenemos $\tilde{y}_t$ para todo t podemos calcular la señal de salida $\mathbf{y} = \mathbf{E}\tilde{\mathbf{y}}$
- Existen algoritmos muy rápidos para calcular la respuesta en frecuencia y su inversa
  $\Rightarrow$ Fast Fourier Transform (depende del valor de t) – no usan una multiplicación de matrices
- Muchas implementaciones de la convolución toman el filtro y la señal de entrada
  $\Rightarrow$ Calculan las respuestas en frecuencias y multiplican el resultado
  $\Rightarrow$ Calculan la inversa para obtener la señal de salida

Señales y Sistemas

Convolución

Respuesta en Frecuencia

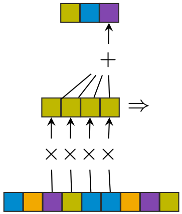
Convoluciones en Imágenes

Convoluciones en Imágenes

- La operación de convolución permite explotar la estructura de los datos
  - $\Rightarrow$ "La estructura de los datos" $\Rightarrow$ Valores contiguos en el vector están relacionados
- Señales temporales: x_i está muy relacionado con x_{i+1}, pero no mucho con x_{i+K} para $K \gg 1$
  - $\Rightarrow$ Noción de frecuencia: Si los valores contiguos cambian mucho, alta frecuencia
  - $\Rightarrow$ Si los valores lejanos cambian mucho, no podemos decir demasiado
- Otro ejemplo son las imágenes: los píxeles contiguos de la imagen están relacionados
  - $\Rightarrow$ Ordenamiento de los píxeles, detección de bordes, etc.

- Sabemos cómo aplicar la convolución a vectores $\Rightarrow$ Vectoricemos la imagen

- Ahora que tenemos el vector, consideremos un filtro y calculemos la convolución



- Relaciona valores contiguos en las columnas, pega un salto cuando una columna termina $\Rightarrow$ No sirve

Convolución en 2-D

- Recordemos la definición general de convolución $\Rightarrow \mathsf{h} * \mathsf{x} = \sum_{k=0}^{K-1} h_k \mathsf{D}_k(\mathsf{x})$

  $\Rightarrow$ El problema es que la noción de desplazamiento $\mathsf{D}_k(\mathsf{x}) = \mathsf{x}(n-k)$ no sirve

- Las imágenes se definen como $\mathsf{x} : \mathcal{M} \times \mathcal{N} \to \mathbb{R}^3$ $\Rightarrow$ Una coordenada vertical y una horizontal $\mathsf{x}(m, n)$
- Hay, entonces, dos dimensiones de desplazamiento $\mathsf{D}_{p,q}(\mathsf{x}) = \mathsf{x}(m-p, n-q)$
- Los filtros también tienen dos coordenadas $\Rightarrow \mathsf{h}(m, n)$ para avanzar en ambas direcciones
- La convolución en imágenes se define como

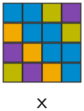
$$\mathsf{y}(m, n) = \sum_{p=0}^{K} \sum_{q=0}^{K'} \mathsf{h}(p, q)\mathsf{x}(m-p, n-q)$$

  $\Rightarrow$ La salida también es una imagen $\Rightarrow$ Tiene dos dimensiones

  $\Rightarrow$ El filtro no tiene por qué ser cuadrado (aunque generalmente lo es) $(K \ll M,\ K' \ll N)$

Convolución en 2-D: Interpretación Gráfica

- La operación de convolución es: combinación lineal, desplazar, combinación lineal, desplazar, ...
- Tenemos dos señales: una es el filtro h y la otra son los datos x



- La convolución implica una combinación lineal y un desplazamiento

Convolución en Imágenes: Observaciones

- La convolución en imágenes es una convolución
  - $\Rightarrow$ Las propiedades que vimos para la convolución en una dimensión valen
  - $\Rightarrow$ Linealidad, aditividad, rellenar con ceros, etc.
- También existe una respuesta en frecuencia para imágenes
  - $\Rightarrow$ También hace el cálculo de la convolución más rápido

Resumen del Capítulo

- A veces, los datos tienen estructuras que se pueden explotar para mejorar el algoritmo
  $\Rightarrow$ Señales en el tiempo, sistemas para procesarlos
- Operación de convolución $\Rightarrow$ Operación lineal que explota la estructura
- Respuesta en frecuencia $\Rightarrow$ Representación alternativa de señales con estructura
  $\Rightarrow$ Permite encontrar una manera rápida y fácil de calcular la convolución
- Convolución en imágenes $\Rightarrow$ Conceptualmente igual que la convolución en el tiempo

Escenas del Próximo Capítulo...

- La convolución es una operación lineal $\Rightarrow$ Pero vimos que es mejor aprender algoritmos no lineales
  $\Rightarrow$ Redes neuronales convolucionales $\Rightarrow$ Reemplazar la operación lineal por una convolución
  $\Rightarrow$ Ganar control sobre la representación $\Rightarrow$ Múltiples features y bancos de filtros
- Pooling $\Rightarrow$ Reducir la dimensionalidad en cada capa construyendo resúmenes
  $\Rightarrow$ Quedarse únicamente con resúmenes regionales $\Rightarrow$ Invarianza local
- Algunas arquitecturas típicas usadas en la práctica