

## Modelos probabilísticos estructurados en DL.

- Los modelos de DL se pueden estudiar desde diferentes perspectivas.
- Los modelos probabilísticos estructurados pretenden describir las distribuciones de probabilidad usando grafos para incluir las interacciones entre variables aleatorias.
- **Grafo**: conjunto de vértices (nodos) conectados a través de bordes (edges).
- También conocidos como modelos gráficos graphical-models .

## Problemas de los modelos no estructurados.

- DL → extensión del aprendizaje de máquina a inteligencia artificial.
  - Se requiere modelar datos de alta dimensión con estructuras complejas.
  - Por ejemplo: clasificación convencional resume toda la entrada en una etiqueta o probabilidad de membresía
- NOTA:** Bajo modelos probabilísticos se busca resolver problemas más complejos: trabajar con múltiples salidas, modelar la estructura de los datos, etc.

## Algunas tareas de interés:

- Estimación de pdf:  $x \in \mathbb{R}^P$ ;  $p(x)$ ?
- Filtrado:  $\tilde{x} \in \mathbb{R}^P$   $\tilde{x} = x + \eta$ ;  $x = ?$
- Completar datos perdidos:  $x \in \mathbb{R}^P$ ;  $x_j = \text{NaN}$   
 $\hat{x}_j = ?$
- Muestreo:  $x \sim p(x)$

**NOTA:** Se busca modelar una distribución para  $P \rightarrow \infty$ .

Ej: Modelar imágenes de  $32 \times 32$  en RGB

$$\# \text{pixels} = 32 \times 32 \times 3 = 3072.$$

Si se restringe  $x_{ijc} \in \{0, 1\}$ ;  $2^{3072}$  combinaciones.

→ Si se tiene un vector aleatorio  $x$  con  $P$  variables discretas con  $K$  posibles valores cada una,  $P(x)$  se puede modelar desde la tabla:

	$x_1^1$	$x_1^2$	$\dots$	$x_1^K$
$x_1^1$		$\dots$		
$x_1^2$		$\dots$		
$\vdots$		$\ddots$		
$x_1^K$		$\dots$		

$$P(x_1^e, x_m^r); e, r \in \{1, 2, \dots, K\}$$
$$j, m \in \{1, 2, \dots, P\}.$$

→ La estimación de  $P(x)$  por tabla no es eficiente  
→ Problemas de memoria, tiempo de proceso, inestabilidad y cantidad de datos requeridos.

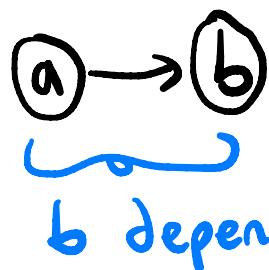
- En  $P(x)$  por tabla se pretende modelar cada posible interacción entre variables
- En la práctica las  $P(x)$  son más simples.
- La mayoría de las variables influyen en otras de forma indirecta.
- Modelos probabilísticos estructurados. proveen solo relaciones directas entre variables aleatorias.

## Grafos para describir la estructura del modelo

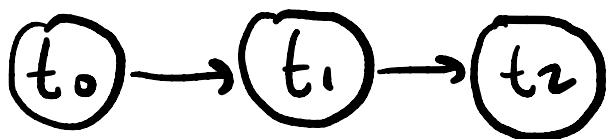
- Objetivo: representar interacciones entre variables aleatorias
- Estructuras básicas:
  - grafos acíclicos dirigidos
  - grafos no dirigidos.

## Modelos dirigidos:

- Directed graphical model → Belief network
  - Bayesian Network.



Se define la distribución sobre b como  $p(b|a)$



$$p(t_0, t_1, t_2) = p(t_0)p(t_1|t_0)p(t_2|t_1)$$

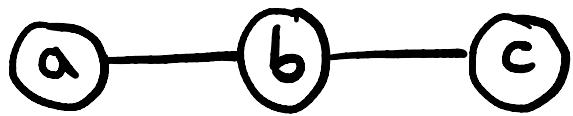
Un modelo gráfico dirigido en x se define como un grafo acíclico dirigido  $G$  cuyos vértices son variables aleatorias y posee un conjunto de probabilidades condicionales locales  $p(x_i | Pa_G(x_i))$ , donde  $Pa_G(x_i)$  relaciona las conexiones en  $x_i$ .

$$p(x) = \prod_i p(x_i | Pa_G(x_i))$$

**NOTA:** Menor número de conexiones en el grafo implica que la distribución se puede modelar con menor parámetros.

### Modelos no dirigidos.

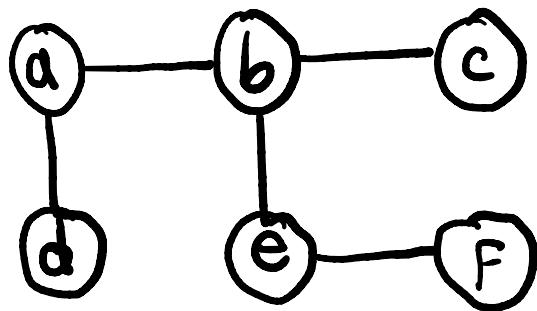
- Conocidos también como Markov Random Fields
- o Redes de Markov.
- Para modelos dirigidos debe ser clara la conexión dirección entre variables aleatorias.
- Si no es clara la dirección o se relacionan en ambas direcciones, se sugiere trabajar con modelos no dirigidos



**NOTA:** las conexiones no se pueden asociar mediante probabilidades condicionales.

- Para cada ciclo  $C$  del grafo (subconjunto de nodos interconectados), un factor  $\phi(C)$  cuantifica la afinidad de las variables.
- Los  $\phi(c) \geq 0$ .
- Así, se puede definir una probabilidad no normalizada:

$$\tilde{p}(x) = \prod_{C \in \mathcal{E}} \phi(C)$$



$$p(a, b, c, d, e, f) = \frac{1}{Z} \phi_{a,b}(a, b) \phi_{b,c}(b, c) \phi_{a,d}(a, d) \phi_{b,e}(b, e) \phi_{e,f}(e, f)$$

## Función de partición

$$p(x) = \frac{1}{Z} \tilde{p}(x)$$

$$Z = \int \tilde{p}(x) dx$$

**NOTA:** Cuando una distribución se define normalizando un producto de potenciales cíclicos se denomina distribución de Gibbs.

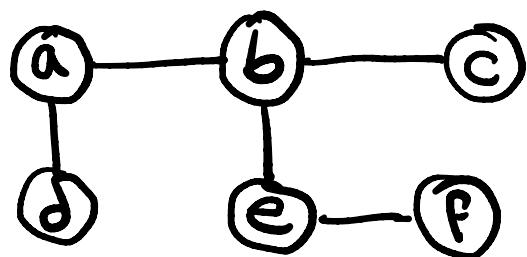
- $Z$  se conoce también como partition function.
- $Z = \int \tilde{p}(x) dx$  generalmente no es tratable, se requieren aprox.
- La escogencia de  $\phi(x)$  es crítica.  
Por ej:  $\phi(x) = x^2$   
 $Z = \int x^2 dx \rightarrow \infty$   
 $\phi(x; \beta) = \exp(-\beta x^2)$   
 $Z = \int \exp(-\beta x^2) dx < \infty \text{ si } \beta \in \mathbb{R}^+$

### Modelos basados en energía.

- Se asume  $\tilde{p}(x) > 0, \forall x$ , mediante:  
$$\tilde{p}(x) = \exp(-E(x)) \quad (*)$$
- $E(x)$ : función de energía.
- Cualquier distribución como en (\*) se conoce como la distribución de Boltzmann (Boltzmann machine)

→ En sus orígenes Boltzmann solo se restringía a variables binarias.

→ Si:  $\phi(a,b)\phi(c,d) = \exp(a)\exp(b) = \exp(a+b)$   
entonces los diferentes términos del grafo no dirigido se vuelven términos de la función de energía



$$\tilde{p}(a,b,c,d,e,f) = \phi(a,b)\phi(b,c)\phi(b,e)\phi(a,d)\phi(e,f)$$

$$E(a,b,c,d,e,f) = E(a,b) + E(a,d) + E(b,c) + E(b,e) + E(e,f)$$

→ Si se trabaja sobre  $\log \tilde{p}(x)$ ; y se optimiza sobre este término, denominado free-energy:

$$\mathcal{F}(x) = -\log \sum_h \exp(-E(x,h))$$

$h$ : variables latentes