

ANÁLISIS DE TÓPICOS: MODELOS PROBABILÍSTICOS

Salvador López Mendoza

Mayo de 2018

PROBLEMAS AL USAR UN TÉRMINO COMO TÓPICO

- Falta poder de expresión.
 - Solamente se pueden representar tópicos muy sencillos o muy generales.
 - No es posible representar tópicos complicados.
- Cobertura incompleta del vocabulario.
 - No es posible capturar las variaciones en el vocabulario (palabras relacionadas, sinónimos).
- Ambigüedad en la identificación de las palabras.
 - El término que define los tópicos puede ser ambiguo. Por ejemplo *estrella*.
la estrella de cine, es una estrella del fútbol.

PROBLEMAS AL USAR UN TÉRMINO COMO TÓPICO

- Falta poder de expresión. **tópico = {varias palabras}**
 - Solamente se pueden representar tópicos muy sencillos o muy generales.
 - No es posible representar tópicos complicados.
- Cobertura incompleta del vocabulario.
 - No es posible capturar las variaciones en el vocabulario (palabras relacionadas, sinónimos).
- Ambigüedad en la identificación de las palabras.
 - El término que define los tópicos puede ser ambiguo. Por ejemplo *estrella*.
la estrella de cine, es una estrella del fútbol.

PROBLEMAS AL USAR UN TÉRMINO COMO TÓPICO

- Falta poder de expresión. **tópico = {varias palabras}**
 - Solamente se pueden representar tópicos muy sencillos o muy generales.
 - No es posible representar tópicos complicados.
- Cobertura incompleta del vocabulario. **mayor peso a las palabras**
 - No es posible capturar las variaciones en el vocabulario (palabras relacionadas, sinónimos).
- Ambigüedad en la identificación de las palabras.
 - El término que define los tópicos puede ser ambiguo. Por ejemplo *estrella*.
la estrella de cine, es una estrella del fútbol.

PROBLEMAS AL USAR UN TÉRMINO COMO TÓPICO

- Falta poder de expresión. **tópico = {varias palabras}**
 - Solamente se pueden representar tópicos muy sencillos o muy generales.
 - No es posible representar tópicos complicados.
- Cobertura incompleta del vocabulario. **mayor peso a las palabras**
 - No es posible capturar las variaciones en el vocabulario (palabras relacionadas, sinónimos).
- Ambigüedad en la identificación de las palabras. **separar palabras ambiguas**
 - El término que define los tópicos puede ser ambiguo. Por ejemplo *estrella*.
la estrella de cine, es una estrella del fútbol.

PROBLEMAS AL USAR UN TÉRMINO COMO TÓPICO

- Falta poder de expresión. **tópico = {varias palabras}**
 - Solamente se pueden representar tópicos muy sencillos o muy generales.
 - No es posible representar tópicos complicados.
- Cobertura incompleta del vocabulario. **mayor peso a las palabras**
 - No es posible capturar las variaciones en el vocabulario (palabras relacionadas, sinónimos).
- Ambigüedad en la identificación de las palabras. **separar palabras ambiguas**
 - El término que define los tópicos puede ser ambiguo. Por ejemplo *estrella*.
la estrella de cine, es una estrella del fútbol.

Modelo probabilístico de tópicos

DISTRIBUCIÓN DE PALABRAS

DISTRIBUCIÓN DE PALABRAS

Definir cada tópico como el conjunto de palabras que se le relacionan.

DISTRIBUCIÓN DE PALABRAS

Definir cada tópico como el conjunto de palabras que se le relacionan.

DEPORTES (θ_1): deportes (0.02), juego (0.01), fútbol (0.004), *jugar* (0.003), *estrella* (0.003), ganar (0.002), *viajar* (0.0005).

DISTRIBUCIÓN DE PALABRAS

Definir cada tópico como el conjunto de palabras que se le relacionan.

DEPORTES (θ_1): deportes (0.02), juego (0.01), fútbol (0.004), *jugar* (0.003), *estrella* (0.003), ganar (0.002), *viajar* (0.0005).

VIAJAR (θ_2): *viajar* (0.05), vuelo (0.004), hotel (0.003), playa (0.002), cultura (0.0004), *jugar* (0.0002).

DISTRIBUCIÓN DE PALABRAS

Definir cada tópico como el conjunto de palabras que se le relacionan.

DEPORTES (θ_1): deportes (0.02), juego (0.01), fútbol (0.004), *jugar* (0.003), *estrella* (0.003), ganar (0.002), *viajar* (0.0005).

VIAJAR (θ_2): *viajar* (0.05), vuelo (0.004), hotel (0.003), playa (0.002), cultura (0.0004), *jugar* (0.0002).

CIENCIA (θ_k): ciencia (0.04), científico (0.03), cohete (0.006), telescopio (0.004), genoma (0.0004), *estrella* (0.0002), genética (0.0001), *viajar* (0.00001).

DISTRIBUCIÓN DE PALABRAS

Definir cada tópico como el conjunto de palabras que se le relacionan.

DEPORTES (θ_1): deportes (0.02), juego (0.01), fútbol (0.004), *jugar* (0.003), *estrella* (0.003), ganar (0.002), *viajar* (0.0005).

VIAJAR (θ_2): *viajar* (0.05), vuelo (0.004), hotel (0.003), playa (0.002), cultura (0.0004), *jugar* (0.0002).

CIENCIA (θ_k): ciencia (0.04), científico (0.03), cohete (0.006), telescopio (0.004), genoma (0.0004), *estrella* (0.0002), genética (0.0001), *viajar* (0.00001).

Propiedad:

DISTRIBUCIÓN DE PALABRAS

Definir cada tópico como el conjunto de palabras que se le relacionan.

DEPORTES (θ_1): deportes (0.02), juego (0.01), fútbol (0.004), *jugar* (0.003), *estrella* (0.003), ganar (0.002), *viajar* (0.0005).

VIAJAR (θ_2): *viajar* (0.05), vuelo (0.004), hotel (0.003), playa (0.002), cultura (0.0004), *jugar* (0.0002).

CIENCIA (θ_k): ciencia (0.04), científico (0.03), cohete (0.006), telescopio (0.004), genoma (0.0004), *estrella* (0.0002), genética (0.0001), *viajar* (0.00001).

Propiedad:

Para un vocabulario $V = \{w_1, w_2, \dots\}$

$$\sum_{w \in V} p(w|\theta_i) = 1$$

MINADO PROBABILÍSTICO DE TÓPICOS

MINADO PROBABILÍSTICO DE TÓPICOS

- Entrada:

MINADO PROBABILÍSTICO DE TÓPICOS

- Entrada:
 - Una colección de N documentos de texto. $C = \{d_1, d_2, \dots, d_N\}$.

MINADO PROBABILÍSTICO DE TÓPICOS

- Entrada:
 - Una colección de N documentos de texto. $C = \{d_1, d_2, \dots, d_N\}$.
 - Un vocabulario. $V = \{w_1, \dots, w_M\}$.

MINADO PROBABILÍSTICO DE TÓPICOS

- Entrada:
 - Una colección de N documentos de texto. $C = \{d_1, d_2, \dots, d_N\}$.
 - Un vocabulario. $V = \{w_1, \dots, w_M\}$.
 - k , la cantidad de tópicos.

MINADO PROBABILÍSTICO DE TÓPICOS

- Entrada:
 - Una colección de N documentos de texto. $C = \{d_1, d_2, \dots, d_N\}$.
 - Un vocabulario. $V = \{w_1, \dots, w_M\}$.
 - k , la cantidad de tópicos.
- Salida:

MINADO PROBABILÍSTICO DE TÓPICOS

- Entrada:
 - Una colección de N documentos de texto. $C = \{d_1, d_2, \dots, d_N\}$.
 - Un vocabulario. $V = \{w_1, \dots, w_M\}$.
 - k , la cantidad de tópicos.
- Salida:
 - k tópicos.

MINADO PROBABILÍSTICO DE TÓPICOS

- Entrada:
 - Una colección de N documentos de texto. $C = \{d_1, d_2, \dots, d_N\}$.
 - Un vocabulario. $V = \{w_1, \dots, w_M\}$.
 - k , la cantidad de tópicos.
- Salida:
 - **k tópicos.**
Cada uno es una distribución de palabras:
 $\{\theta_1, \dots, \theta_k\}$, con $\sum_{w \in V} p(w|\theta_i) = 1$

MINADO PROBABILÍSTICO DE TÓPICOS

- Entrada:
 - Una colección de N documentos de texto. $C = \{d_1, d_2, \dots, d_N\}$.
 - Un vocabulario. $V = \{w_1, \dots, w_M\}$.
 - k , la cantidad de tópicos.
- Salida:
 - **k tópicos.**
Cada uno es una distribución de palabras:
 $\{\theta_1, \dots, \theta_k\}$, con $\sum_{w \in V} p(w|\theta_i) = 1$
 - Una cobertura de tópicos en cada documento d_i .
 $d_i : \{\pi_{i1}, \dots, \pi_{ik}\}$

MINADO PROBABILÍSTICO DE TÓPICOS

- Entrada:

- Una colección de N documentos de texto. $C = \{d_1, d_2, \dots, d_N\}$.
- Un vocabulario. $V = \{w_1, \dots, w_M\}$.
- k , la cantidad de tópicos.

- Salida:

- k tópicos.

Cada uno es una distribución de palabras:

$\{\theta_1, \dots, \theta_k\}$, con $\sum_{w \in V} p(w|\theta_i) = 1$

- Una cobertura de tópicos en cada documento d_i .

$d_i : \{\pi_{i1}, \dots, \pi_{ik}\}$

π_{ij} es la probabilidad de que el tópico θ_j esté cubierto en el documento d_i .

MINADO PROBABILÍSTICO DE TÓPICOS

- Entrada:

- Una colección de N documentos de texto. $C = \{d_1, d_2, \dots, d_N\}$.
- Un vocabulario. $V = \{w_1, \dots, w_M\}$.
- k , la cantidad de tópicos.

- Salida:

- k tópicos.

Cada uno es una distribución de palabras:

$\{\theta_1, \dots, \theta_k\}$, con $\sum_{w \in V} p(w|\theta_i) = 1$

- Una cobertura de tópicos en cada documento d_i .

$d_i : \{\pi_{i1}, \dots, \pi_{ik}\}$

π_{ij} es la probabilidad de que el tópico θ_j esté cubierto en el documento d_i .

$$\sum_{j=1}^k \pi_{ij} = 1$$

FUNCIONAMIENTO

FUNCIONAMIENTO

Dada la colección de documentos (C), el vocabulario (V) y el valor de k .

FUNCIONAMIENTO

Dada la colección de documentos (C), el vocabulario (V) y el valor de k .
Se obtiene el conjunto de tópicos ($\{\theta_1, \dots, \theta_k\}$).

FUNCIONAMIENTO

Dada la colección de documentos (C), el vocabulario (V) y el valor de k .

Se obtiene el conjunto de tópicos ($\{\theta_1, \dots, \theta_k\}$).

Se puede calcular la cobertura de cada tópico en cada documento d_i :
 $\{\pi_{i1}, \dots, \pi_{ik}\}$.

FUNCIONAMIENTO

Dada la colección de documentos (C), el vocabulario (V) y el valor de k .

Se obtiene el conjunto de tópicos ($\{\theta_1, \dots, \theta_k\}$).

Se puede calcular la cobertura de cada tópico en cada documento d_i :
 $\{\pi_{i1}, \dots, \pi_{ik}\}$.

Cada documento está asociado al tópico que tenga mayor porcentaje de cobertura.

FUNCIONAMIENTO

Dada la colección de documentos (C), el vocabulario (V) y el valor de k .

Se obtiene el conjunto de tópicos ($\{\theta_1, \dots, \theta_k\}$).

Se puede calcular la cobertura de cada tópico en cada documento d_i :
 $\{\pi_{i1}, \dots, \pi_{ik}\}$.

Cada documento está asociado al tópico que tenga mayor porcentaje de cobertura.

¿Hay una única clasificación?

FUNCIONAMIENTO

Dada la colección de documentos (C), el vocabulario (V) y el valor de k .

Se obtiene el conjunto de tópicos ($\{\theta_1, \dots, \theta_k\}$).

Se puede calcular la cobertura de cada tópico en cada documento d_i :
 $\{\pi_{i1}, \dots, \pi_{ik}\}$.

Cada documento está asociado al tópico que tenga mayor porcentaje de cobertura.

¿Hay una única clasificación?

Depende de varios factores:

FUNCIONAMIENTO

Dada la colección de documentos (C), el vocabulario (V) y el valor de k .

Se obtiene el conjunto de tópicos ($\{\theta_1, \dots, \theta_k\}$).

Se puede calcular la cobertura de cada tópico en cada documento d_i :
 $\{\pi_{i1}, \dots, \pi_{ik}\}$.

Cada documento está asociado al tópico que tenga mayor porcentaje de cobertura.

¿Hay una única clasificación?

Depende de varios factores:

Cantidad de tópicos (k), la función de selección de tópicos.

MODELO GENERATIVO

MODELO GENERATIVO

Se tienen diferentes escenarios, dependiendo de los parámetros.

MODELO GENERATIVO

Se tienen diferentes escenarios, dependiendo de los parámetros.

Modelado de generación de datos: $P(\text{Datos}|\text{Modelo}, \Lambda)$

MODELO GENERATIVO

Se tienen diferentes escenarios, dependiendo de los parámetros.

Modelado de generación de datos: $P(\text{Datos}|\text{Modelo}, \Lambda)$

Es un modelo probabilístico.

MODELO GENERATIVO

Se tienen diferentes escenarios, dependiendo de los parámetros.

Modelado de generación de datos: $P(\text{Datos}|\text{Modelo}, \Lambda)$

Es un modelo probabilístico.

$$\Lambda = (\{\theta_1, \dots, \theta_k\}, \{\pi_{11}, \dots, \pi_{1k}\}, \dots, \{\pi_{N1}, \dots, \pi_{Nk}\})$$

MODELO GENERATIVO

Se tienen diferentes escenarios, dependiendo de los parámetros.

Modelado de generación de datos: $P(\text{Datos}|\text{Modelo}, \Lambda)$

Es un modelo probabilístico.

$$\Lambda = (\{\theta_1, \dots, \theta_k\}, \{\pi_{11}, \dots, \pi_{1k}\}, \dots, \{\pi_{N1}, \dots, \pi_{Nk}\})$$

Para un conjunto de datos en particular, se debe inferir el conjunto de parámetros más adecuado (Λ^*), tal que

$$\Lambda^* = \operatorname{argmax}_{\Lambda} P(\text{Datos}|\text{Modelo}, \Lambda)$$

MODELO GENERATIVO

Se tienen diferentes escenarios, dependiendo de los parámetros.

Modelado de generación de datos: $P(\text{Datos}|\text{Modelo}, \Lambda)$

Es un modelo probabilístico.

$$\Lambda = (\{\theta_1, \dots, \theta_k\}, \{\pi_{11}, \dots, \pi_{1k}\}, \dots, \{\pi_{N1}, \dots, \pi_{Nk}\})$$

Para un conjunto de datos en particular, se debe inferir el conjunto de parámetros más adecuado (Λ^*), tal que

$$\Lambda^* = \operatorname{argmax}_{\Lambda} P(\text{Datos}|\text{Modelo}, \Lambda)$$

Λ^* es el **conocimiento** que se debe extraer (*minar*) del problema.

MODELO GENERATIVO

Se tienen diferentes escenarios, dependiendo de los parámetros.

Modelado de generación de datos: $P(\text{Datos}|\text{Modelo}, \Lambda)$

Es un modelo probabilístico.

$$\Lambda = (\{\theta_1, \dots, \theta_k\}, \{\pi_{11}, \dots, \pi_{1k}\}, \dots, \{\pi_{N1}, \dots, \pi_{Nk}\})$$

Para un conjunto de datos en particular, se debe inferir el conjunto de parámetros más adecuado (Λ^*), tal que

$$\Lambda^* = \operatorname{argmax}_{\Lambda} P(\text{Datos}|\text{Modelo}, \Lambda)$$

Λ^* es el **conocimiento** que se debe extraer (*minar*) del problema.

Se pueden hacer ajustes al modelo para descubrir conocimientos distintos.

MODELO ESTADÍSTICO DEL LENGUAJE (LM)

MODELO ESTADÍSTICO DEL LENGUAJE (LM)

¿Qué es un modelo estadístico del lenguaje?

MODELO ESTADÍSTICO DEL LENGUAJE (LM)

¿Qué es un modelo estadístico del lenguaje?

- Distribución de probabilidad sobre las secuencias de palabras.

MODELO ESTADÍSTICO DEL LENGUAJE (LM)

¿Qué es un modelo estadístico del lenguaje?

- Distribución de probabilidad sobre las secuencias de palabras.
 $p(\text{"Hoy es martes"}) \approx 0.001$

MODELO ESTADÍSTICO DEL LENGUAJE (LM)

¿Qué es un modelo estadístico del lenguaje?

- Distribución de probabilidad sobre las secuencias de palabras.

$$p(\text{"Hoy es martes"}) \approx 0.001$$

$$p(\text{"Martes es hoy"}) \approx 0.000000001$$

$$p(\text{"La derivada parcial"}) \approx 0.00001$$

MODELO ESTADÍSTICO DEL LENGUAJE (LM)

¿Qué es un modelo estadístico del lenguaje?

- Distribución de probabilidad sobre las secuencias de palabras.
 $p(\text{"Hoy es martes"}) \approx 0.001$
 $p(\text{"Martes es hoy"}) \approx 0.000000001$
 $p(\text{"La derivada parcial"}) \approx 0.00001$
- Es dependiente del contexto.

MODELO ESTADÍSTICO DEL LENGUAJE (LM)

¿Qué es un modelo estadístico del lenguaje?

- Distribución de probabilidad sobre las secuencias de palabras.
 $p(\text{"Hoy es martes"}) \approx 0.001$
 $p(\text{"Martes es hoy"}) \approx 0.000000001$
 $p(\text{"La derivada parcial"}) \approx 0.00001$
- Es dependiente del contexto.
- Es un mecanismo probabilístico para *generar* texto.

MODELO ESTADÍSTICO DEL LENGUAJE (LM)

¿Qué es un modelo estadístico del lenguaje?

- Distribución de probabilidad sobre las secuencias de palabras.
 $p(\text{"Hoy es martes"}) \approx 0.001$
 $p(\text{"Martes es hoy"}) \approx 0.000000001$
 $p(\text{"La derivada parcial"}) \approx 0.00001$
- Es dependiente del contexto.
- Es un mecanismo probabilístico para *generar* texto.
Es un modelo generativo.

MODELO SENCILLO: LM DE UNIGRAMA

MODELO SENCILLO: LM DE UNIGRAMA

- El texto se genera al ir generando cada palabra **independientemente**.

MODELO SENCILLO: LM DE UNIGRAMA

- El texto se genera al ir generando cada palabra **independientemente**.
- $p(w_1 w_2 \dots w_n) = p(w_1)p(w_2) \dots p(w_n)$

MODELO SENCILLO: LM DE UNIGRAMA

- El texto se genera al ir generando cada palabra **independientemente**.
- $p(w_1 w_2 \dots w_n) = p(w_1)p(w_2) \dots p(w_n)$
- Parámetros del modelo: $\{p(w_i)\}$

MODELO SENCILLO: LM DE UNIGRAMA

- El texto se genera al ir generando cada palabra **independientemente**.
- $p(w_1 w_2 \dots w_n) = p(w_1)p(w_2) \dots p(w_n)$
- Parámetros del modelo: $\{p(w_i)\}$
Se tiene que $p(w_1) + \dots + p(w_N) = 1$.

MODELO SENCILLO: LM DE UNIGRAMA

- El texto se genera al ir generando cada palabra **independientemente**.
- $p(w_1 w_2 \dots w_n) = p(w_1)p(w_2) \dots p(w_n)$
- Parámetros del modelo: $\{p(w_i)\}$
Se tiene que $p(w_1) + \dots + p(w_N) = 1$.
 N es el tamaño del vocabulario.

MODELO SENCILLO: LM DE UNIGRAMA

- El texto se genera al ir generando cada palabra **independientemente**.
- $p(w_1 w_2 \dots w_n) = p(w_1)p(w_2) \dots p(w_n)$
- Parámetros del modelo: $\{p(w_i)\}$
Se tiene que $p(w_1) + \dots + p(w_N) = 1$.
 N es el tamaño del vocabulario.
- El texto es una muestra sobre el conjunto de posibles combinaciones de palabras de acuerdo a una distribución de las palabras.

MODELO SENCILLO: LM DE UNIGRAMA

- El texto se genera al ir generando cada palabra **independientemente**.
- $p(w_1 w_2 \dots w_n) = p(w_1)p(w_2) \dots p(w_n)$
- Parámetros del modelo: $\{p(w_i)\}$
Se tiene que $p(w_1) + \dots + p(w_N) = 1$.
 N es el tamaño del vocabulario.
- El texto es una muestra sobre el conjunto de posibles combinaciones de palabras de acuerdo a una distribución de las palabras.
 $p(\text{"Hoy es martes"}) = p(\text{"Hoy"})p(\text{"es"})p(\text{"martes"})$

GENERACIÓN DE TEXTOS CON LM DE UNIGRAMA

GENERACIÓN DE TEXTOS CON LM DE UNIGRAMA

Dado un documento d

GENERACIÓN DE TEXTOS CON LM DE UNIGRAMA

Dado un documento d

y un conjunto de tópicos, $\{\theta_1, \dots, \theta_k\}$,

GENERACIÓN DE TEXTOS CON LM DE UNIGRAMA

Dado un documento d

y un conjunto de tópicos, $\{\theta_1, \dots, \theta_k\}$,

se aplica el LM de unigrama al documento d con cada tópico θ_i ($p(d|\theta_i)$)

GENERACIÓN DE TEXTOS CON LM DE UNIGRAMA

Dado un documento d

y un conjunto de tópicos, $\{\theta_1, \dots, \theta_k\}$,

se aplica el LM de unigrama al documento d con cada tópico θ_i ($p(d|\theta_i)$)

El documento se refiere al tópico con el mayor valor.

ESTIMACIÓN DEL MODELO LM DE UNIGRAMA

ESTIMACIÓN DEL MODELO LM DE UNIGRAMA

¿Qué tan bueno es el cálculo obtenido?

ESTIMACIÓN DEL MODELO LM DE UNIGRAMA

¿Qupe tan bueno es el cálculo obtenido?

Depende del total de palabras del texto y la frecuencia con la que aparecen.

ESTIMACIÓN DEL MODELO LM DE UNIGRAMA

¿Qué tan bueno es el cálculo obtenido?

Depende del total de palabras del texto y la frecuencia con la que aparecen.

Al fijar estos valores se obtienen valores para cada documento.

ESTIMACIÓN DEL MODELO LM DE UNIGRAMA

¿Qupe tan bueno es el cálculo obtenido?

Depende del total de palabras del texto y la frecuencia con la que aparecen.

Al fijar estos valores se obtienen valores para cada documento.

¿Es el **mejor** valor que se puede esperar?

ESTIMACIÓN DEL MODELO LM DE UNIGRAMA

¿Qupe tan bueno es el cálculo obtenido?

Depende del total de palabras del texto y la frecuencia con la que aparecen.

Al fijar estos valores se obtienen valores para cada documento.

¿Es el **mejor** valor que se puede esperar?

¿Cómo se define *mejor*?

MÁXIMA SIMILITUD VS. BAYESIANA

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene una valor de similitud mayor?

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene un valor de similitud mayor?
Dado X , ¿cuál θ maximiza $p(X, \theta)$?

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene un valor de similitud mayor?
Dado X , ¿cuál θ maximiza $p(X, \theta)$?
 - Problema. Se trabaja con muestras pequeñas.

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene un valor de similitud mayor?
Dado X , ¿cuál θ maximiza $p(X, \theta)$?
 - Problema. Se trabaja con muestras pequeñas.
- Estimación bayesiana.

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene un valor de similitud mayor?
Dado X , ¿cuál θ maximiza $p(X, \theta)$?
 - Problema. Se trabaja con muestras pequeñas.
- Estimación bayesiana.
 - *Mejor* quiere decir ser consistente con el conocimiento *previo* y explica bien los datos.

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene un valor de similitud mayor?
Dado X , ¿cuál θ maximiza $p(X, \theta)$?
 - Problema. Se trabaja con muestras pequeñas.
- Estimación bayesiana.
 - *Mejor* quiere decir ser consistente con el conocimiento *previo* y explica bien los datos.
 - Problema. ¿Cómo se define *previo*?

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene un valor de similitud mayor?
Dado X , ¿cuál θ maximiza $p(X, \theta)$?
 - Problema. Se trabaja con muestras pequeñas.
- Estimación bayesiana.
 - *Mejor* quiere decir ser consistente con el conocimiento *previo* y explica bien los datos.
 - Problema. ¿Cómo se define *previo*?
 - Estimación/inferencia bayesiana ($f(\theta)$):

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene un valor de similitud mayor?
Dado X , ¿cuál θ maximiza $p(X, \theta)$?
 - Problema. Se trabaja con muestras pequeñas.
- Estimación bayesiana.
 - *Mejor* quiere decir ser consistente con el conocimiento *previo* y explica bien los datos.
 - Problema. ¿Cómo se define *previo*?
 - Estimación/inferencia bayesiana ($f(\theta)$):
Debe definir un *previo*: $p(\theta)$

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene un valor de similitud mayor?
Dado X , ¿cuál θ maximiza $p(X, \theta)$?
 - Problema. Se trabaja con muestras pequeñas.
- Estimación bayesiana.
 - *Mejor* quiere decir ser consistente con el conocimiento *previo* y explica bien los datos.
 - Problema. ¿Cómo se define *previo*?
 - Estimación/inferencia bayesiana ($f(\theta)$):
Debe definir un *previo*: $p(\theta)$
Distribución *posterior*.

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene un valor de similitud mayor?
Dado X , ¿cuál θ maximiza $p(X, \theta)$?
 - Problema. Se trabaja con muestras pequeñas.
- Estimación bayesiana.
 - *Mejor* quiere decir ser consistente con el conocimiento *previo* y explica bien los datos.
 - Problema. ¿Cómo se define *previo*?
 - Estimación/inferencia bayesiana ($f(\theta)$):
Debe definir un *previo*: $p(\theta)$
Distribución *posterior*.
 $p(\theta|X)$ es como $p(X|\theta)p(\theta)$

MÁXIMA SIMILITUD VS. BAYESIANA

- Estimación por máxima similitud.
 - *Mejor* quiere decir *la similitud entre los datos alcanza su máximo*.
 - Función de similitud: $p(X|\theta)$
Dado el valor de θ , ¿cuál X tiene un valor de similitud mayor?
Dado X , ¿cuál θ maximiza $p(X, \theta)$?
 - Problema. Se trabaja con muestras pequeñas.
- Estimación bayesiana.
 - *Mejor* quiere decir ser consistente con el conocimiento *previo* y explica bien los datos.
 - Problema. ¿Cómo se define *previo*?
 - Estimación/inferencia bayesiana ($f(\theta)$):
Debe definir un *previo*: $p(\theta)$
Distribución *posterior*.
 $p(\theta|X)$ es como $p(X|\theta)p(\theta)$
Permite inferir cualquier *valor derivado* a partir de θ .