Redes Bayesianas: Clasificación supervisada (III)

Aritz Pérez[1] Borja Calvo[2]

Basque Center for Applied Mathematics

UPV/EHU

Donostia, Febrero de 2015

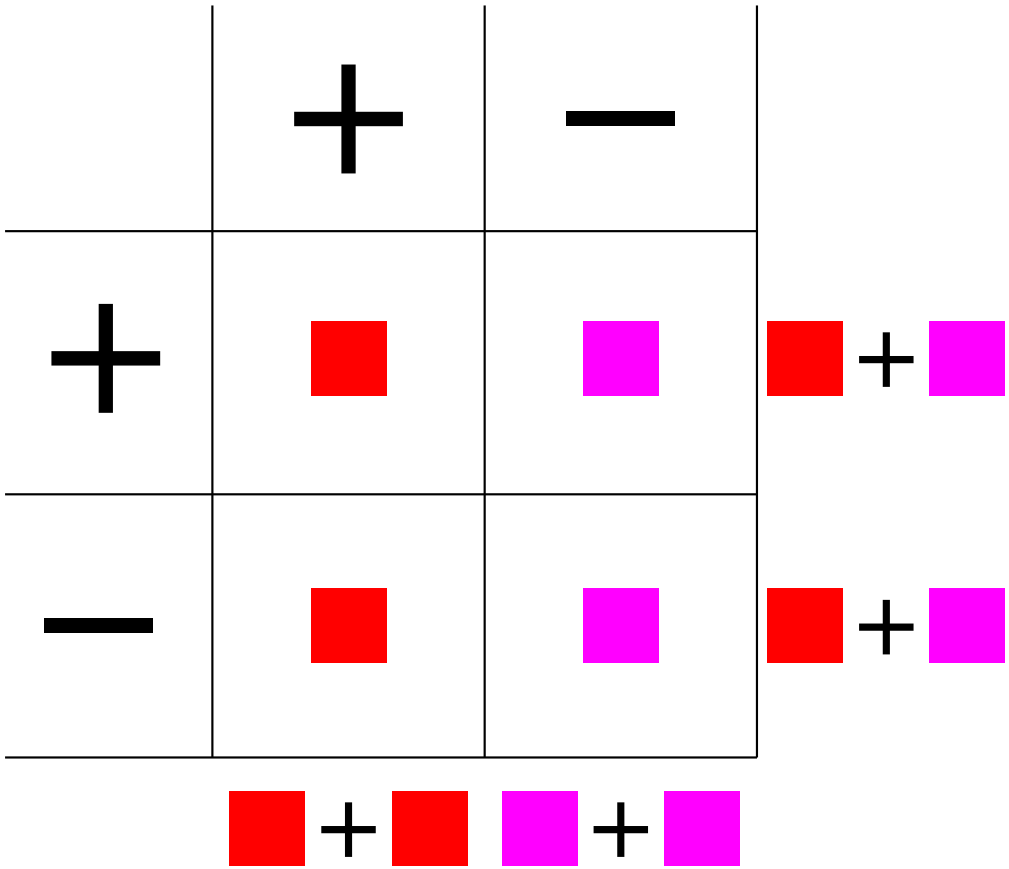
Bibliografía

- K.P. Murphy (2012). Machine Learning: **A Probabilistic Perspective**. The MIT Press.

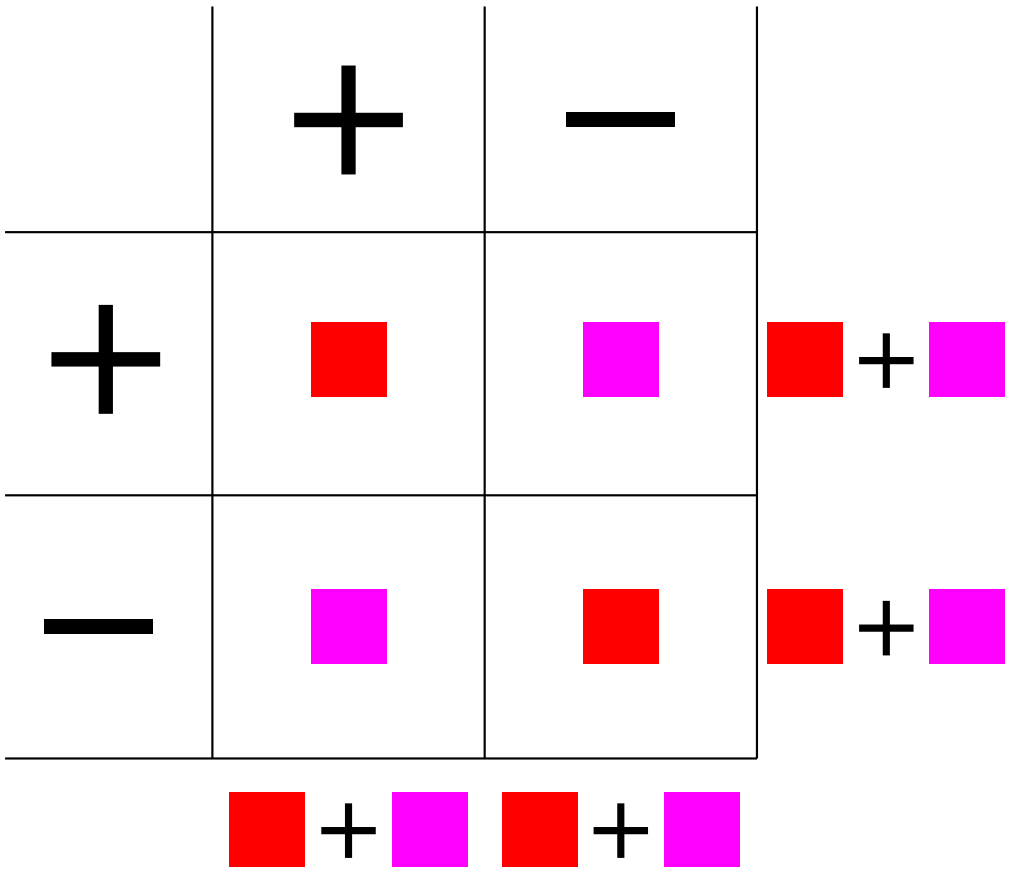
Estructuras sesgadas para la clasificación

- Pocos **parámetros** y mucha información **discriminativa**
- Dependencias importantes $d(X_i; C|\mathbf{X}_S)$
- Dependencias algo menos importantes $d(\mathbf{X}_A; \mathbf{X}_B|C)$

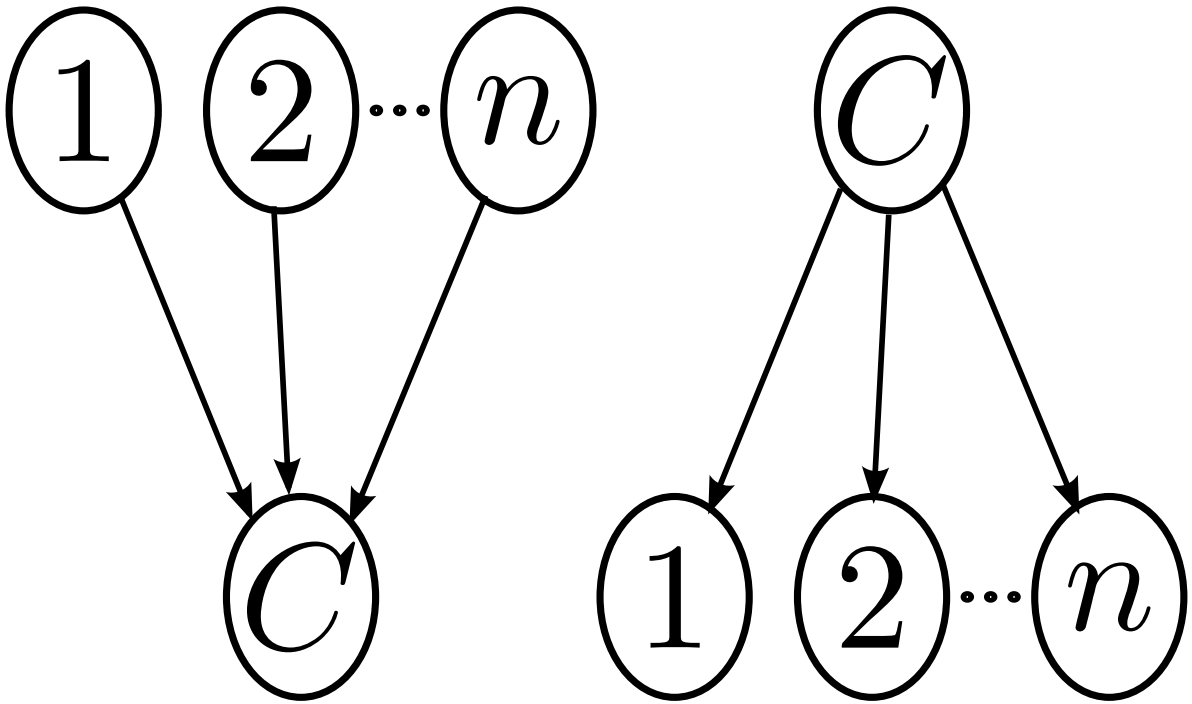
Información discriminativa y dependencias importantes



Información discriminativa y dependencias importantes

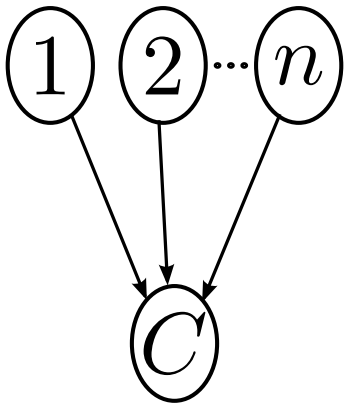


Estructuras sesgadas para la clasificación



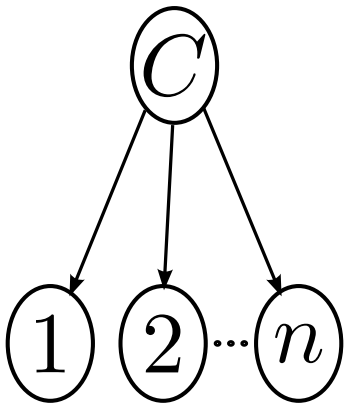
- Qué estructura es preferible para la clasificación?

Muchos parámetros



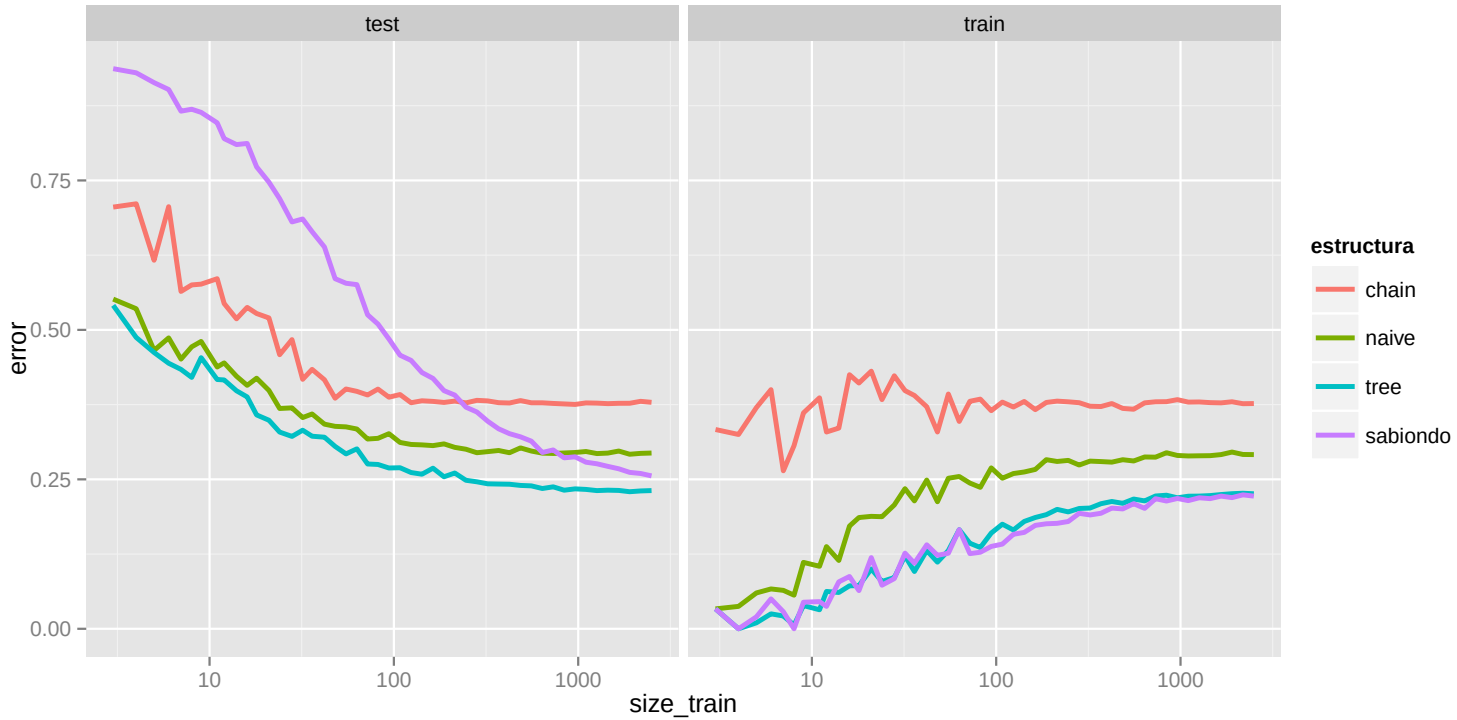
- Modela **todas** las dependencias importantes
- Con suficientes datos modela $p(c|\mathbf{x})$ de forma **perfecta**
- Número de parámetros **exponencial** en n
- Riesgo alto de **sobreajuste**

Naïve Bayes



- Suposición: $i(X_i; X_j|C)$
- Modela las dependencias **más importantes** $\{d(X_i; C|\mathbf{X}_S)\}$
- Número de parametros **lineal** en n
- **Poco** riesgo de **sobreajuste**

Estructuras sesgadas para la clasificación



Naïve Bayes aumentado a árbol

- **Romper con la suposición** del naïve Bayes
- Permitir que cada predictora tenga un padre (además de la clase): **árbol**
- Algoritmo de filtrado **eficiente y óptimo**
- Existen otras generalizaciones, e.g. limitar el máximo **número de padres**