

Redes Bayesianas: Clasificación supervisada (II)

Aritz Pérez¹ Borja Calvo²

Basque Center for Applied Mathematics

UPV/EHU

Donostia, Febrero de 2015

Bibliografía

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

Información discriminativa

- Información que contienen las **predictoras** acerca de la clase
- Desigualdad de Fano: el **error de Bayes** esta relacionado con la **entropía condicionada** $H(C|\mathbf{X})$
- Cuanta **menos incertidumbre** tengamos sobre C cuando conocemos \mathbf{X} el **problema** es **más fácil**
- **Minimizar** la entropía condicionada $H_M(C|\mathbf{X})$ o **maximizar** la información mutua $I_M(C; \mathbf{X})$

Dependencias importantes

- Interesa modelar bien $p(C|\mathbf{x})$ pero no $p(\mathbf{X})$
- Evitar las independencias $i(C; \mathbf{X}_A | \mathbf{X}_{V \setminus A})$
- En menor grado evitar las independencias $i(\mathbf{X}_A; \mathbf{X}_B | C)$
- Introducir las independencias $i(\mathbf{X}_A; \mathbf{X}_B | \mathbf{X}_C)$

Aprendizaje

- Aprendizaje **estructural**
- Modelar la **información** discriminativa
- Selección de **variables**
- Aproximaciones: de **filtrado** y de **envoltura**

Aproximación de filtrado (filter)

- Maximización de una **función objetivo**: LL, BIC,...
- Basados en test de **independencia condicionada**

Problemas de la aproximación de filtrado

- Problema de la verosimilitud y de sus versiones penalizadas
- Conforme n aumenta $p(\mathbf{x})$ tiene **más importancia** que $p(c|\mathbf{x})$:
 - $p(\mathbf{x})$ se hace **más pequeño** y $p(c|\mathbf{x})$ se **mantiene**
 - $H(\mathbf{X}) \gg H(C|\mathbf{X})$
 - La **parte menos importante** para la **clasificación** es más importante para la función de evaluación
- Verosimilitud **condicionada**

Aproximación de envoltura (wrapper)

- Basada en la función que queremos optimizar: **error** de clasificación
- Función **desconocida**
- Emplea un **estimador** de dicha función: validación cruzada
- Escoger uno con **mínima varianza**: repeticiones y estratificación

Problemas de la aproximación de envoltura

- Función de evaluación **no descomponible**
- Suele ser más **costosa de evaluar**

La maldición de la dimensionalidad

- Muchas **variables** en comparación con el número de **casos**
- Riesgo de **sobreajuste**
- Variables **redundantes**
- Variables **irrelevantes**
- Solución: **selección** o **proyección** de variables

Selección de variables

- Reducción del **ruido** y de la varianza
- Aproximaciones de filtrado: **test de independencia** condicional ($X_i; C|\emptyset$)
- Aproximación de envoltura: autoajuste en base al **error**
- **Manto de Markov**