

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Reducción de la dimensionalidad

Eduardo Morales, Hugo Jair Escalante

INAOE

Contenido

1 Introducción

2 Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

3 Generación de atributos

PCA

LDA

GP

AE

4 Discusión

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Reducción de a dimensionalidad

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

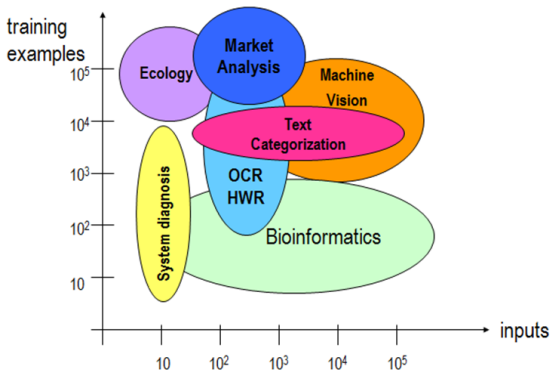
AE

Discusión

En algunos problemas de reconocimiento de patrones, nos enfrentamos a conjuntos de datos de alta dimensionalidad. Ejemplo: clasificación de textos:

- Los documentos se representan por su BoW, esto es vectores con tantos elementos como palabras en el vocabulario.
- Para bases de datos con 1000-10,000 documentos se pueden llegar a tener vocabularios de varias decenas/cientos de miles de elementos.

Reducción de a dimensionalidad



I. Guyon, et al. **Feature Extraction: Foundations and Applications**, Springer 2006.

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Reducción de a dimensionalidad

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Problemas de la alta dimensionalidad en los datos:

- Costo de procesamiento y almacenamiento.
- Atributos relevantes e irrelevantes.
- Calculo de distancias no es sencillo.
- *Data sparsity*.
- La maldición de la dimensionalidad.

Reducción de a dimensionalidad

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

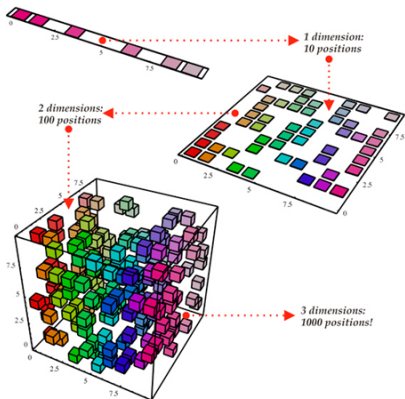
Problemas de la alta dimensionalidad en los datos:

- Costo de procesamiento y almacenamiento.
- Atributos relevantes e irrelevantes.
- Calculo de distancias no es sencillo.
- *Data sparsity*.
- **La maldición de la dimensionalidad.**

Reducción de a dimensionalidad

El número de posiciones escala de manera exponencial con la dimensionalidad del problema.

Necesitamos un número exponencial de ejemplos de entrenamiento para cubrir todas las posiciones.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Reducción de a dimensionalidad

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

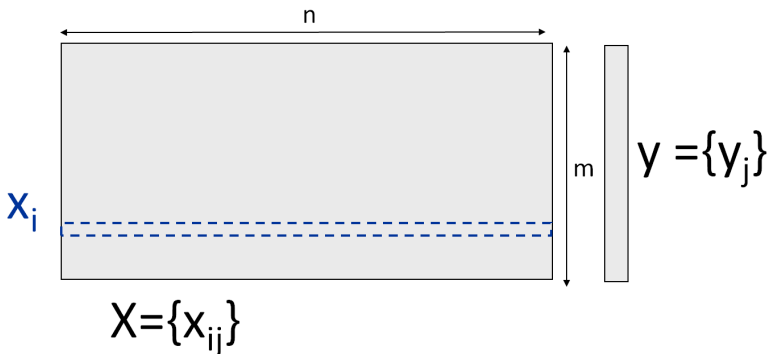
Discusión

Dos soluciones principales:

- **Selección de atributos:** Seleccionar un subconjunto de atributos.
- **Generación de atributos (re-parametrización/transformación):** Mapear los atributos originales a un nuevo espacio de menor dimensión.

Selección de atributos

- **Problema:** Encontrar un subconjunto de atributos que son más útiles para clasificación.
- **Objetivos:** Eliminar atributos irrelevantes/ruidosos; seleccionar/mantener atributos relevantes; reducir la dimensionalidad.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

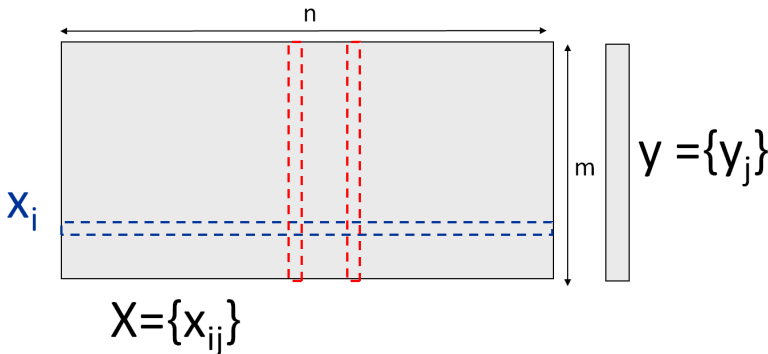
GP

AE

Discusión

Selección de atributos

- **Problema:** Encontrar un subconjunto de atributos que son más útiles para clasificación.
- **Objetivos:** Eliminar atributos irrelevantes/ruidosos; seleccionar/mantener atributos relevantes; reducir la dimensionalidad.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

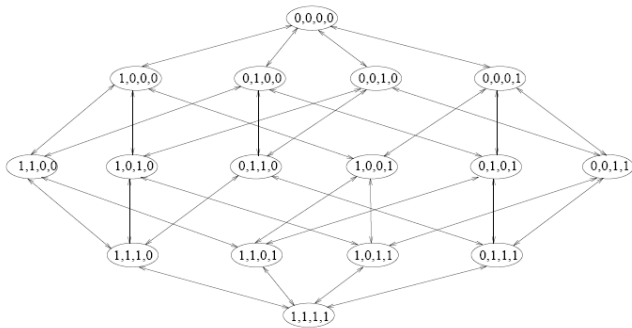
Generación de atributos

PCA
LDA
GP
AE

Discusión

Selección de atributos

Para un problema con d -atributos, cuántos subconjuntos existen?



Problema computacionalmente costoso!

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Selección de atributos

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

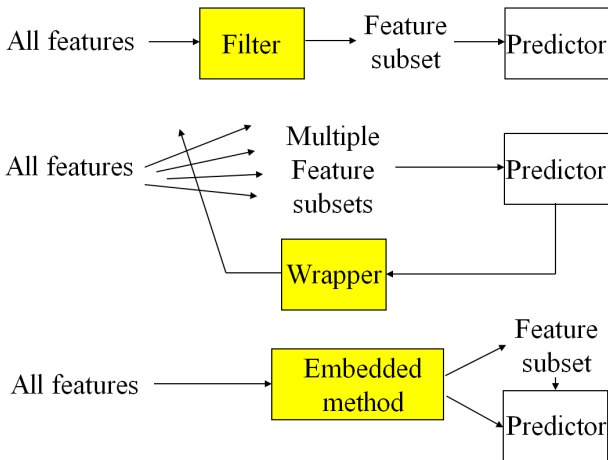
AE

Discusión

Tres enfoques principales:

- **Filtros:** Evalúan la importancia de los atributos usando métodos que son independientes del modelo de clasificación.
- **Wrappers:** Evalúan la importancia de subconjuntos de atributos usando un modelo de clasificación (se adopta una estrategia de búsqueda).
- **Embedded:** Toman ventaja de la naturaleza del modelo de clasificación considerado.

Selección de atributos



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Selección de atributos - Isabelle Guyon

Filters

Methods:

- **Criterion:** Measure feature/feature subset “relevance”
- **Search:** Usually order features (individual feature ranking or nested subsets of features)
- **Assessment:** Use statistical tests

Results:

- Are (relatively) robust against overfitting
- May fail to select the most “useful” features

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Selección de atributos - Isabelle Guyon

Wrappers

Methods:

- Criterion: A risk functional
- Search: Search the space of feature subsets
- Assessment: Use cross-validation

Results:

- Can in principle find the most “useful” features, but
- Are prone to overfitting

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Selección de atributos - Isabelle Guyon

Embedded Methods

Methods:

- Criterion: A risk functional
- Search: **Search guided by the learning process**
- Assessment: Use cross-validation

Results:

- Similar to wrappers, but
- Less computationally expensive
- Less prone to overfitting

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

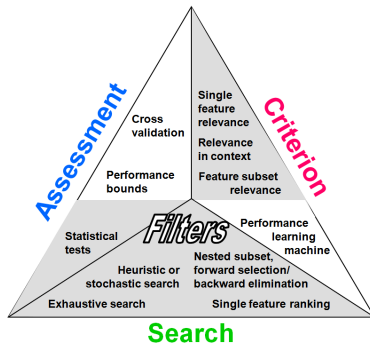
GP

AE

Discusión

Selección de atributos - Isabelle Guyon

Three “Ingredients”



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

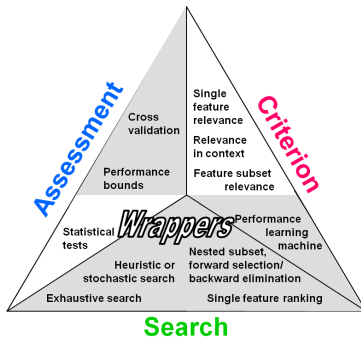
GP

AE

Discusión

Selección de atributos - Isabelle Guyon

Three “Ingredients”



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

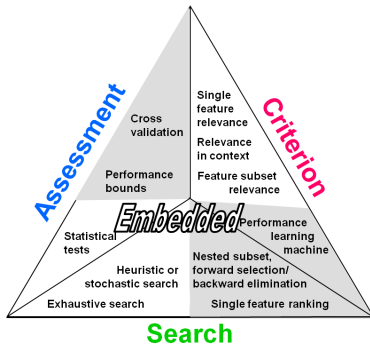
GP

AE

Discusión

Selección de atributos - Isabelle Guyon

Three “Ingredients”



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

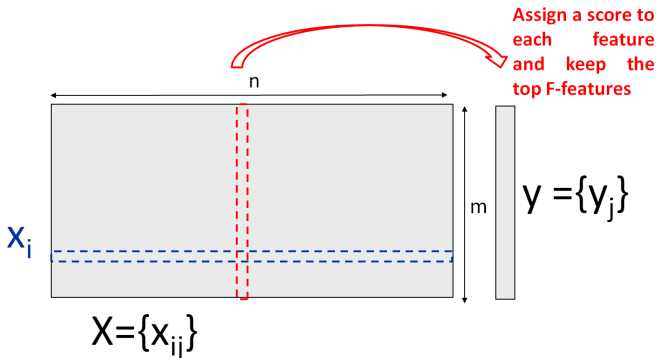
GP

AE

Discusión

Filtros para selección de atributos

Idea: Asignar un score a cada uno de los atributos, y seleccionar los mejores (algunos métodos son multi-variables).



Slide taken from I. Guyon. **Feature and Model Selection**. Machine Learning Summer School, Ile de Re, France, 2008.

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA
LDA
GP
AE

Discusión

Filtros para selección de atributos

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA
LDA
GP
AE

Discusión

Correlación: Forma sencilla de calcular la correlación lineal entre dos variables. $\rho(X_i, Y)$ Nos indica la correlación entre el atributo X_i y la clase Y .

$$\rho(X_i, Y) = \frac{COV(X_i, Y)}{\sigma_{X_i} \sigma_Y}$$

con:

$$COV(X_i, Y) = E[(X_i - E[X_i])(Y - E[Y])]$$

$$\sigma_W = \sqrt{E([W - E[W]]^2)}$$

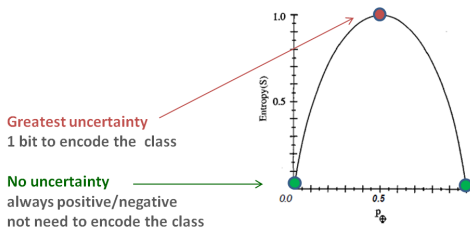
Relación con información mutua.

Filtros para selección de atributos

Ganancia de información (IG). Mide qué tan bien un atributo separa los ejemplos de entrenamiento en dos clases.

Se basa en el concepto de **entropía**, que caracteriza la impureza de una colección arbitraria de ejemplos.

- La entropía especifica el número mínimo de bits de información necesaria para codificar un elemento arbitrario del conjunto de datos.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Filtros para selección de atributos

Ganancia de información (IG). Mide qué tan bien un atributo separa los ejemplos de entrenamiento en dos clases.

Comunmente se mantienen aquellos atributos con IG mayor a 0.

$$IG(X_i, \mathbf{Y}) = - \sum_{i=1}^K Pr(Y_i) \log(Pr(Y_i)) + \dots$$

$$Pr(X_i) \sum_{i=1}^K Pr(Y_i|X_i) \log(Pr(Y_i|X_i)) + \dots$$

$$Pr(\bar{X}_i) \sum_{i=1}^K Pr(Y_i|\bar{X}_i) \log(Pr(Y_i|\bar{X}_i))$$

La IG de un atributo mide la reducción esperada de la entropía causada por la partición de ejemplos usando dicho atributo.

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

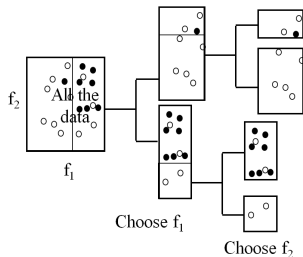
GP

AE

Discusión

Filtros para selección de atributos

Ganancia de información (IG). Mide qué tan bien un atributo separa los ejemplos de entrenamiento en dos clases.



At each step, choose the feature that "reduces entropy" most. Work towards "node purity".

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Filtros para selección de atributos

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

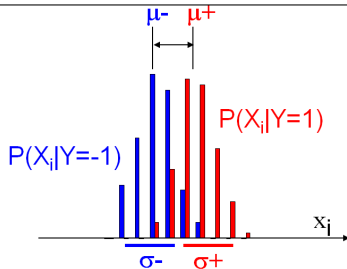
LDA

GP

AE

Discusión

T-test



- Normally distributed classes, equal variance σ^2 unknown; estimated from data as σ^2_{within} .
- Null hypothesis $H_0: \mu^+ = \mu^-$
- T statistic: If H_0 is true,

$$t = (\mu^+ - \mu^-) / (\sigma_{\text{within}} \sqrt{1/m^+ + 1/m^-}) \sim \text{Student}(m^+ + m^- - 2 \text{ d.f.})$$

Otros criterios:

Method	X	Y	Comments					
Name	Formula	B	M	C	B	M	C	
Bayesian accuracy	Eq. 3.1	+	s	+	s			Theoretically the golden standard, rescaled Bayesian relevance Eq. 3.2.
Balanced accuracy	Eq. 3.4	+	s	+	s			Average of sensitivity and specificity; used for unbalanced dataset, same as AUC for binary targets.
Bi-normal separation	Eq. 3.5	+	s	+	s			Used in information retrieval.
F-measure	Eq. 3.7	+	s	+	s			Harmonic of recall and precision, popular in information retrieval.
Odds ratio	Eq. 3.6	+	s	+	s			Popular in information retrieval.
Means separation	Eq. 3.10	+	i	+	+			Based on two class means, related to Fisher's criterion.
T-statistics	Eq. 3.11	+	i	+	+			Based also on the means separation.
Pearson correlation	Eq. 3.9	+	i	+	+	i	+	Linear correlation, significance test Eq. 3.12, or a permutation test.
Group correlation	Eq. 3.13	+	i	+	+	i	+	Pearson's coefficient for subset of features.
χ^2	Eq. 3.8	+	s	+	s			Results depend on the number of samples m .
Relief	Eq. 3.15	+	s	+	+	s	+	Family of methods, the formula is for a simplified version ReliefX, captures local correlations and feature interactions.
Separability Split Value	Eq. 3.41	+	s	+	+	s		Decision tree index.
Kolmogorov distance	Eq. 3.16	+	s	+	+	s	+	Difference between joint and product probabilities.
Bayesian measure	Eq. 3.16	+	s	+	+	s	+	Same as Vajda entropy Eq. 3.23 and Gini Eq. 3.39.
Kullback-Leibler divergence	Eq. 3.20	+	s	+	+	s	+	Equivalent to mutual information.
Jeffreys-Matusita distance	Eq. 3.22	+	s	+	+	s	+	Rarely used but worth trying.
Value Difference Metric	Eq. 3.22	+	s	+	s			Used for symbolic data in similarity-based methods, and symbolic feature-feature correlations.
Mutual Information	Eq. 3.29	+	s	+	+	s	+	Equivalent to information gain Eq. 3.30.
Information Gain Ratio	Eq. 3.32	+	s	+	+	s	+	Information gain divided by feature entropy, stable evaluation.
Symmetrical Uncertainty	Eq. 3.35	+	s	+	+	s	+	Low bias for multivalued features.
J-measure	Eq. 3.36	+	s	+	+	s	+	Measures information provided by a logical rule.
Weight of evidence	Eq. 3.37	+	s	+	+	s	+	So far rarely used.
MDL	Eq. 3.38	+	s	+	s			Low bias for multivalued features.

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

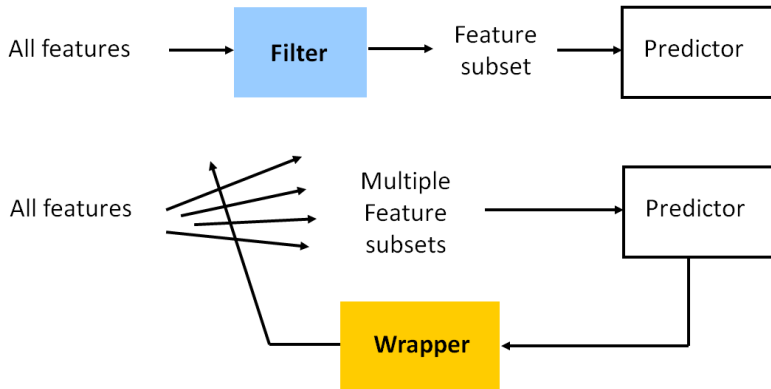
LDA

GP

AE

Discusión

Filtros vs. Wrappers



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

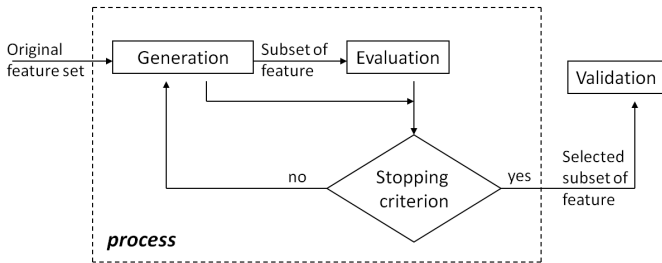
GP

AE

Discusión

Wrappers para selección de atributos

Diagrama general de un método tipo wrapper.



- Generation = select feature subset candidate.
- Evaluation = compute relevancy value of the subset.
- Stopping criterion = determine whether subset is relevant.
- Validation = verify subset validity.

From M. Dash and H. Liu. <http://www.comp.nus.edu.sg/~wongszec/group10.ppt>

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

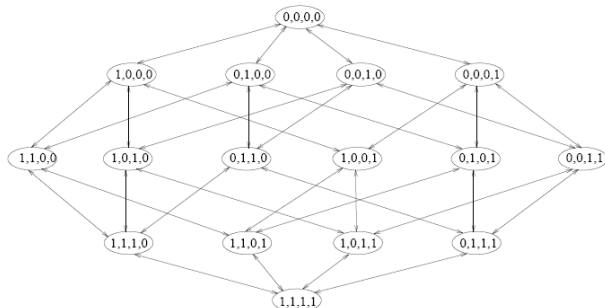
AE

Discusión

Wrappers para selección de atributos

Variantes:

- Estrategia de búsqueda:** Sequential Forward Selection (SFS). Sequential Backward Elimination (SBS). Beam search: keep k best path at each step. Floating search (SFFS and SBFS): Alternate between SFS and SBS as long as we find better subsets than those of the same size obtained so far. Extensive search (simulated annealing, genetic algorithms, exhaustive search).



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Wrappers para selección de atributos

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

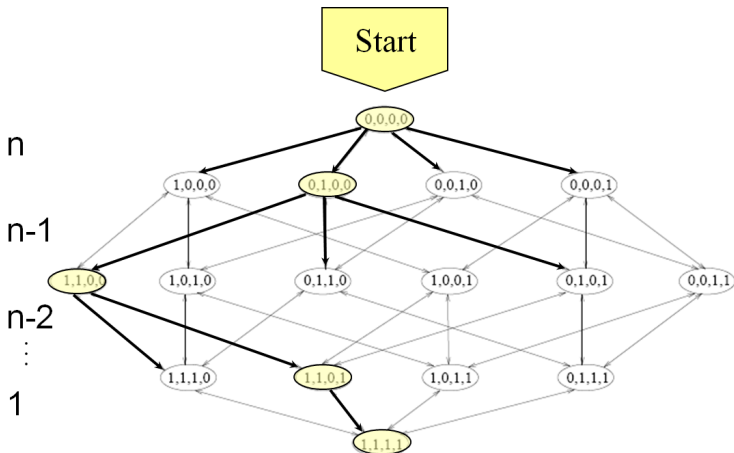
AE

Discusión

Variantes:

- **Criterio:** Generalmente son guiados por el desempeño de clasificación, medido por una función perdida.
- **Muestreo:** Se suele usar validación cruzada u otras estrategias de evaluación para evitar el sobre ajuste.
- **Criterio de paro:** Número máximo de iteraciones, *early stopping*, convergencia.

Forward selection (SFS)



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

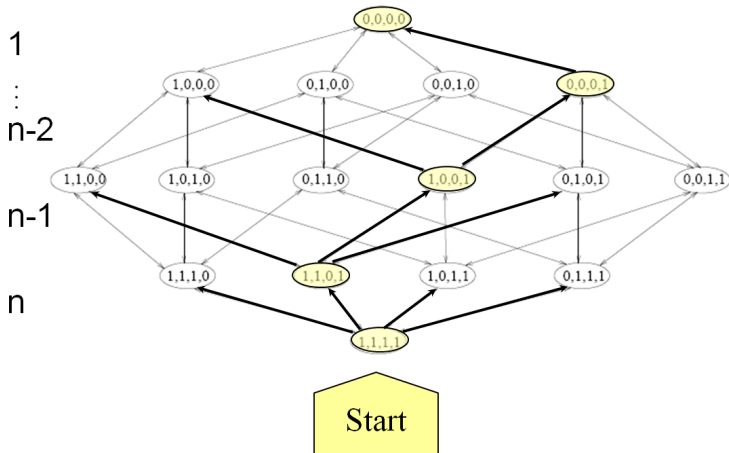
LDA

GP

AE

Discusión

Backward elimination (SBS)



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

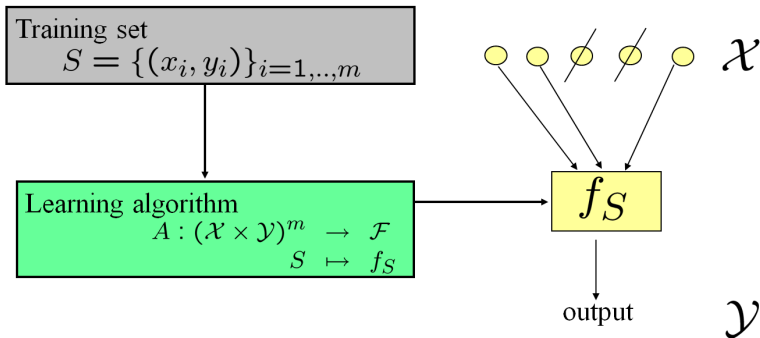
GP

AE

Discusión

Embedded methods

Definition: an embedded feature selection method is a *machine learning algorithm* that returns a model using a limited number of features.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

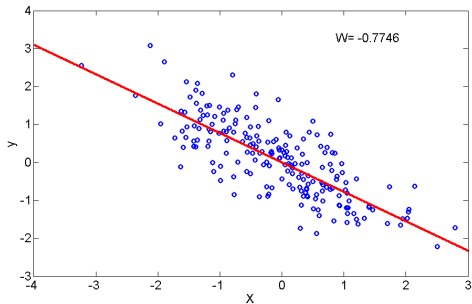
AE

Discusión

Embedded methods

Idea: sacar provecho del modelo o proceso de aprendizaje para identificar variables relevantes.

- Regresión lineal.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Embedded methods

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA
LDA
GP
AE

Discusión

Idea: sacar provecho del modelo o proceso de aprendizaje para identificar variables relevantes.

- Ejemplo: SVM
- Sabiendo que:

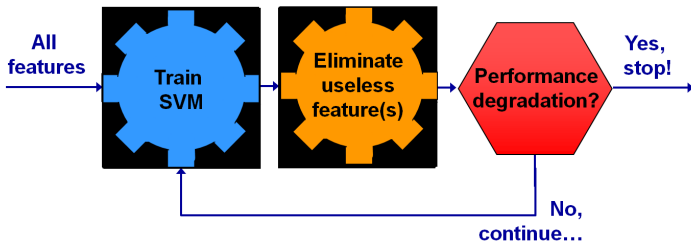
$$\mathbf{w} = \sum_k \alpha_k y_k \mathbf{x}_k \quad (1)$$

- Se usa a \mathbf{w} como criterio para dar importancia a los atributos

Embedded methods

Idea: sacar provecho del modelo o proceso de aprendizaje para identificar variables relevantes.

- SVM.



Recursive Feature Elimination (RFE) SVM. *Guyon-Weston, 2000. US patent 7,117,188*

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

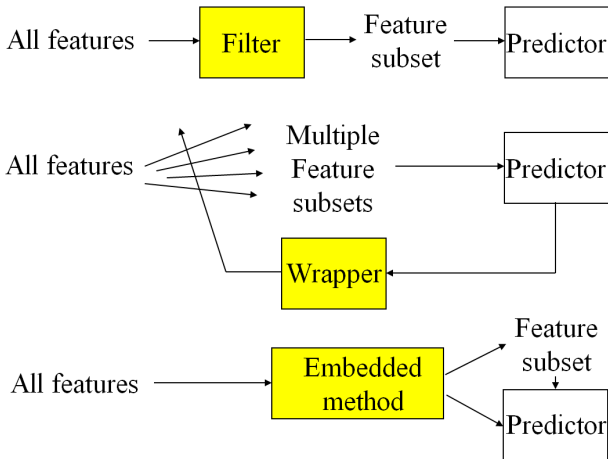
LDA

GP

AE

Discusión

Selección de atributos



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Selección de atributos

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión



Feature Extraction, Foundations and Applications

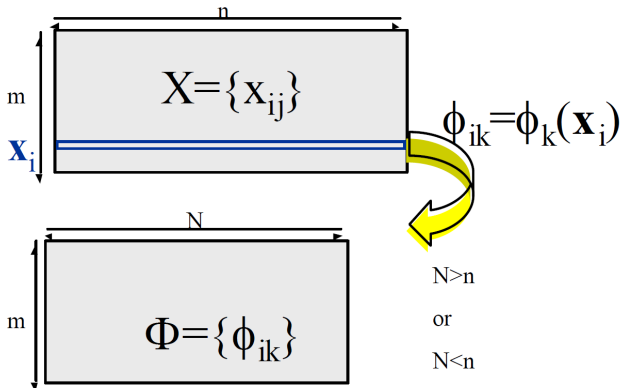
I. Guyon et al, Eds.

Springer, 2006.

<http://clopinet.com/fextract-book>

Generación de atributos

Mapear los atributos originales a un nuevo espacio de menor dimensión.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Generación de atributos

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

- Cómo definir el mapeo $\Phi(\mathbf{x})$?
- Métodos más usados: PCA, LDA, AutoEncoders, Factorización de matrices, etc.

PCA: Análisis de Componentes Principales

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

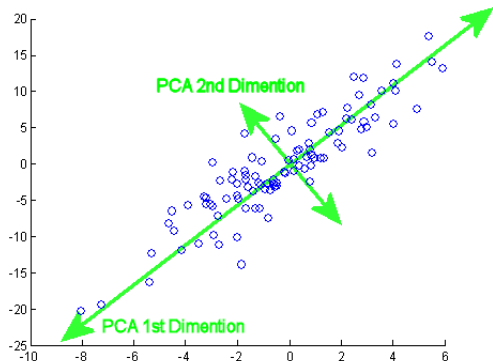
Discusión

PCA: Herramienta estándar en análisis de datos

- Simple
- No paramétrico
- Extrae información relevante a partir de conjuntos de datos confusos
- Provee forma de reducir un conjunto de datos complejo a otro con dimensión menor
- Revela estructuras simplificadas (algunas veces ocultas)
- Permite remover ruido, información no relevante.

PCA: Análisis de Componentes Principales

Idea: Conceptualmente, PCA es un método que encuentra un conjunto de bases que maximizan la varianza de los datos originales y que son ortogonales entre si. Encuentra las direcciones de mayor variación.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

PCA: Análisis de Componentes Principales

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Receta: Sea $\mathcal{X} = \{(\mathbf{x}_i)\}_{1,\dots,N}$ nuestro conjunto de datos de entrenamiento¹ con $\mathbf{x}_i \in \mathbb{R}^d$.

- Centrar los datos: A cada columna \mathbf{X}_j de \mathbf{X} restamos la media de \mathbf{X}_j .
- Calcular la matriz de covarianza:

$$\mathbf{C}_X = \frac{1}{N} \mathbf{X} \mathbf{X}^T$$

- Calcula los eigenvectores \mathbf{v} de \mathbf{C}_X :

$$\mathbf{C}_X \mathbf{v} = \lambda \mathbf{v}$$

- La matriz \mathbf{v} es el conjunto de componentes principales.

¹Ojo: PCA es un método no supervisado.

PCA: Análisis de Componentes Principales

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

- Se pueden proyectar nuevos datos \mathbf{X}^* al espacio de componentes principales mediante:

$$\mathbf{P}^* = \mathbf{v}^T \mathbf{X}^*$$

- Se pueden reconstruir los datos originales:

$$\hat{\mathbf{X}}^* = \mathbf{P}^* \mathbf{v}^T$$

PCA: Para qué sirve?

Cómo se ven los componentes principales?



“eigenfaces”

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

PCA: Para qué sirve?

Cómo se ven los componentes principales?

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

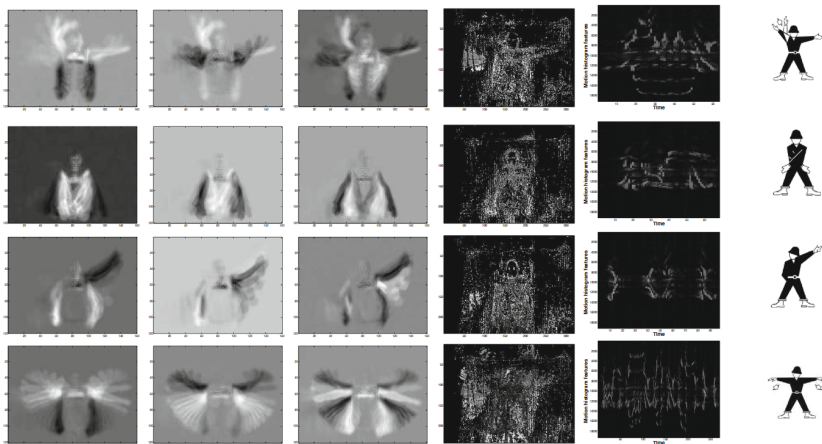
PCA

LDA

GP

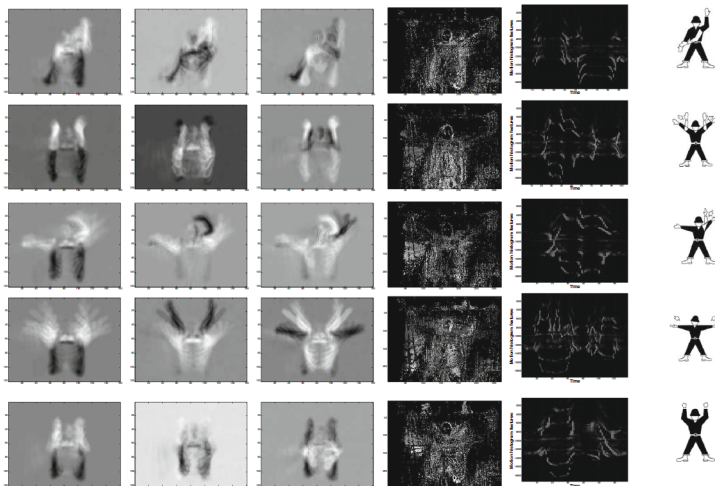
AE

Discusión



PCA: Para qué sirve?

Cómo se ven los componentes principales?



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

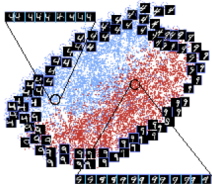
Discusión

PCA: Para qué sirve?

Aplicación 1: Reducción de la dimensionalidad de los datos.

Con PCA obtenemos un conjunto de componentes que nos permiten mapear los datos a otro espacio. Los componentes están ordenados de acuerdo a su importancia para explicar los datos, así, es posible proyectar los datos originales en solo unos cuantos componentes principales.

$$\mathbf{P}^* = \mathbf{v}_{1:k}^T \mathbf{X}^*$$



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

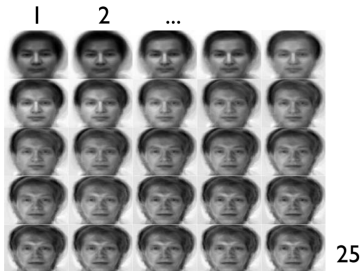
Discusión

PCA: Para qué sirve?

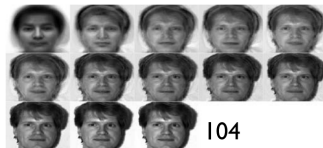
Aplicación 2: Remover ruido, información no relevante de lo datos. La idea es obtener los componentes principales y posteriormente reconstruir los datos usando solo un número pequeño de componentes k ($k \ll d$).

$$\hat{\mathbf{X}}^* = \mathbf{P}^* \mathbf{v}_{1, \dots, k}^T$$

Reconstruction using the first 25 components
(eigenfaces), one at a time

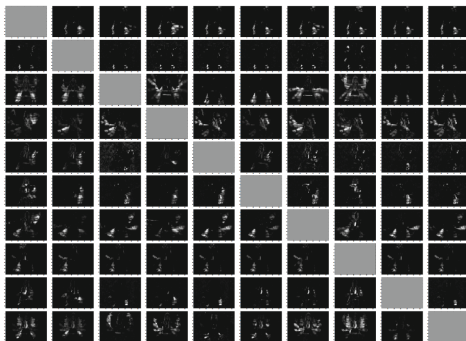


Same, but adding 8 PCA
components at each step



PCA: Para qué sirve?

Aplicación 3: Como clasificador. Se sabe que los componentes principales minimizan el error de reconstrucción de los objetos originales. Entonces, se puede aprender un modelo PCA por cada clase y clasificar nuevas instancias asignándolas a la clase con el menor error de reconstrucción.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

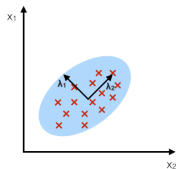
LDA: Análisis de discriminador lineal

PCA es una herramienta que provee transformaciones lineales que capturan la varianza de los datos, sin embargo, es un método no supervisado.

LDA (*Linear discriminant analysis*) es una variante supervisada que nos permite generar atributos tratando de maximizar la discriminatividad.

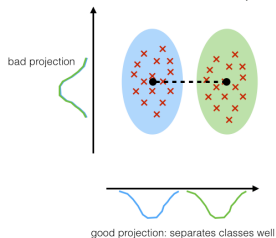
PCA:

component axes that maximize the variance



LDA:

maximizing the component axes for class-separation



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

LDA: Análisis de discriminador lineal

En LDA (para reducción de la dimensionalidad), buscamos la proyección de los datos que nos maximiza la separación entre clases.

$$\mathbf{y} = \mathbf{w}^T \mathbf{x}$$

Queremos encontrar \mathbf{w} que maximiza dicha separación. Fisher planteó el siguiente criterio:

$$J(\mathbf{w}) = \frac{(c_- - c_+)^2}{s_+^2 + s_-^2}$$

con

$$c_+ = \mathbf{w}^T \mathbf{c}_+; c_- = \mathbf{w}^T \mathbf{c}_-$$

$$s_+^2 = \sum_{\{i|y_i=+1\}} (y_i - c_+)^2; s_-^2 = \sum_{\{i|y_i=-1\}} (y_i - c_-)^2;$$

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

LDA: Análisis de discriminador lineal

Haciendo explícita la dependencia con \mathbf{w} , tenemos:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

donde \mathbf{S}_B es la matriz de covarianza entre clases:

$$\mathbf{S}_B = (\mathbf{c}_- - \mathbf{c}_+)(\mathbf{c}_- - \mathbf{c}_+)^T$$

y \mathbf{S}_W es la matriz de covarianza intra clase:

$$\mathbf{S}_W = \sum_{\{i|y_i=+1\}} (\mathbf{x}_i - \mathbf{c}_+)(\mathbf{x}_i - \mathbf{c}_+)^T + \sum_{\{i|y_i=-1\}} (\mathbf{x}_i - \mathbf{c}_-)(\mathbf{x}_i - \mathbf{c}_-)^T$$

La solución está dada por:

$$\mathbf{w} \approx \mathbf{S}_W^{-1}(\mathbf{c}_- - \mathbf{c}_+)$$

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

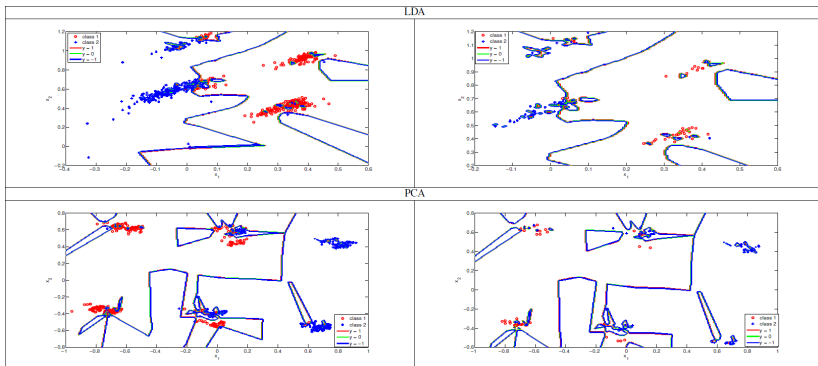
GP

AE

Discusión

LDA: Análisis de discriminador lineal

Proyección de datos con LDA:



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Programación genética como extractor de atributos

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

- Una alternativa a PCA y LDA que trabaja de forma intuitiva es la generación de atributos mediante programación genética.
- La idea para generar un atributo es combinar, mediante ciertas operaciones, atributos del espacio original, e.g.,:

$$\mathbf{w}_i = \Phi_i(\mathbf{X}_S)$$

Donde \mathbf{X}_S es un subconjunto de las columnas de \mathbf{X} y Φ_i es una función que combina dichas columnas.

- El problema consiste en encontrar $\Phi_{1,\dots,k}$

Programación genética

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

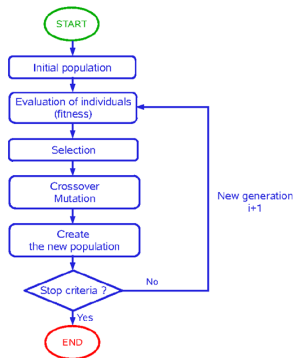
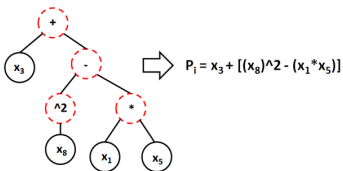
GP

AE

Discusión

Qué es programación genética?

GP: standard EA with tree-based representation



Programación genética

Programación genética para generar atributos

La idea es muy simple: codificar atributos como arboles, y dejar al programa genético que determine los mejores atributos.

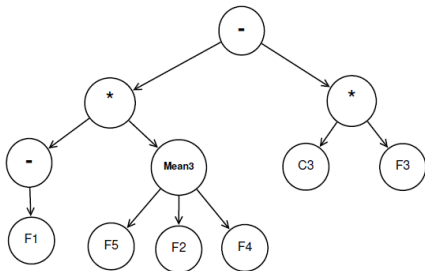


Fig. 2: A feature representation (tree-feature). $\mathcal{F}_y = (-F_1 * \frac{F_5 + F_2 + F_4}{3}) - (C_3 * F_3)$

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Programación genética

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

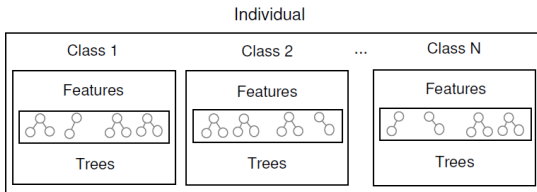
LDA

GP

AE

Discusión

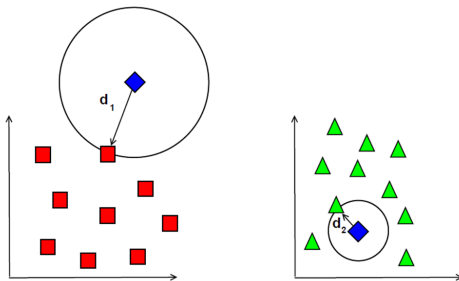
Programación genética para generar atributos
La idea es muy simple: codificar atributos como arboles, y dejar al programa genético que determine los mejores atributos.



Programación genética

Programación genética para generar atributos

La idea es muy simple: codificar atributos como arboles, y dejar al programa genético que determine los mejores atributos.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

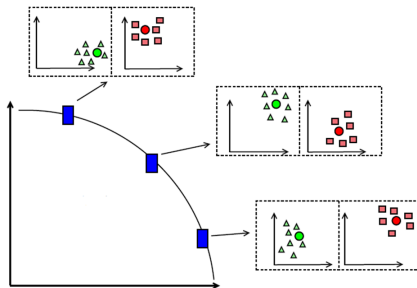
AE

Discusión

Programación genética

Programación genética para generar atributos

La idea es muy simple: codificar atributos como arboles, y dejar al programa genético que determine los mejores atributos.



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Programación genética

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

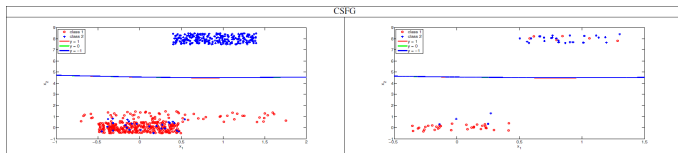
LDA

GP

AE

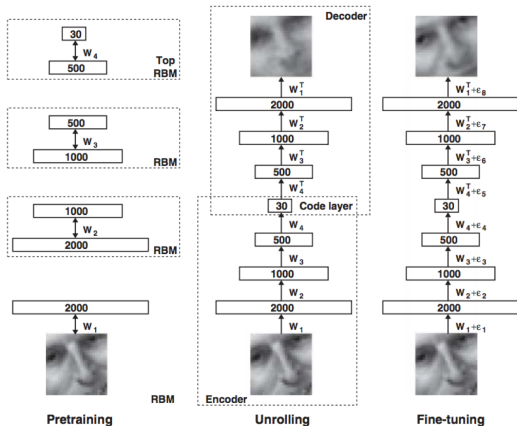
Discusión

Resultados y comparación de espacios generados.



Reprisal: Autoencoders

Los auto encoders pueden verse/usarse como reductores de dimensionalidad:



Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

Comentarios finales

Introducción

Selección de atributos

Filtros para selección de atributos

Wrappers para selección de atributos

Métodos embebidos

Generación de atributos

PCA

LDA

GP

AE

Discusión

- La dimensionalidad puede llegar a ser un problema serio en aprendizaje computacional, cuando ésta es “grande”.
- La maldición de la dimensionalidad afecta mayormente a cierto tipo de métodos.
- Existen dos alternativas: seleccionar o generar atributos.
- Ambas estrategias son útiles y se pueden combinar, además que se les puede dar otros usos.
- En general es un arte, seleccionar el mejor método para un problema en particular.