



Métricas de desempeño para Maquinas de Aprendizaje

Juliana León, Karen Troiano, Stefano Di Colli

November 18, 2014

Índice



Índice



Introducción



Métricas para medir desempeño

- Introducción al problema
- Descripción

Índice





Métricas

- Accuracy
- Lift
- Presion/Recall
 - F1
 - Matthews correlation coefficient
- Matriz de Confusión
 - Sensibility
 - Specificity
- ROC - Curva
- Coeficiente Silhouette

Accuracy



Descripción

- Veracidad: Errores Sistemáticos
- Presión: Errores Random

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Aplicaciones

Aprendizaje de maquinas donde se cuantan con etiquetas

- Supervisado
- No supervisado

Lift



Descripción

- Reglas de asociaciones
- Una parte de la data

Aplicaciones

- Data Mining
- Marketing PONER TABLA



Descripción

- Presion: Predicción de la clase positiva

$$PPV = \frac{TP}{TP + FP}$$

- Recall: Predicción de la clase negativa

$$PPV = \frac{TN}{TN + FN}$$

F1 score



Descripción

- Presion/Recall

$$F1 = 2 \frac{Presion * Recall}{Presion + Recall}$$

Matthews correlation coefficient

- Coeficiente de correlación entre lo observado y lo que se predijo
- Salida : $[-1, 1]$
 - -1 : Predicción completamente opuesta
 - 0: No es mejor que random
 - 1: Predicción Perfecta

Matriz de Confusión



Descripción

Organización de los datos para análisis

	Salida Positiva	Salida Negativa
Real Positiva	TP	FN
Real Negativa	FP	TN

Métricas obtenidas

- Sensitivity:

$$TPR = \frac{TP}{TP + FN}$$

- Specificity:

$$TNR = \frac{TN}{FP + TN}$$

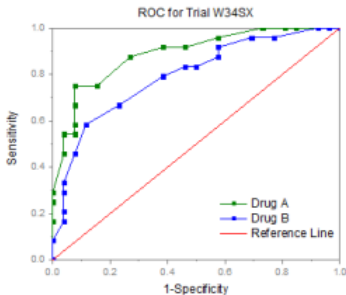
- No datos desbalanceados

ROC



Descripción

- Graficar Sensitivity x Specificity
- Datos desbalanceados
- Medicina y Biología



Coeficiente de Silhouette



Descripción

- Es un coeficiente usado para validar Clusters
- Por cada valor existe un coeficiente Silhouette que representa la clase

Índice



Resumen



- Variedad de metricas
- Adaptivilidad
- Accuracy no es suficiente
- ROC y Matriz de confusion
- Sin los datos etiquetados en un problema mas completo



PREGUNTAS