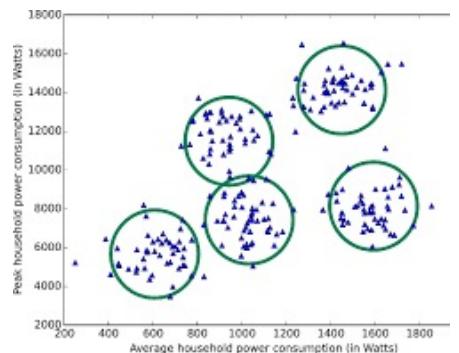
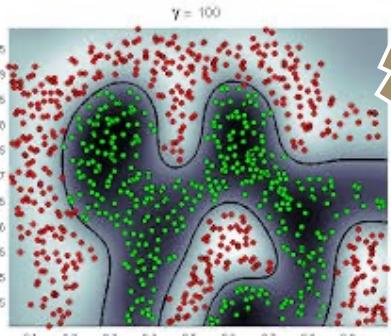




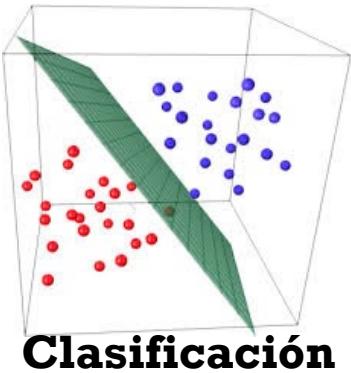
Aprendizaje  
automático



Aprendizaje  
no supervisado



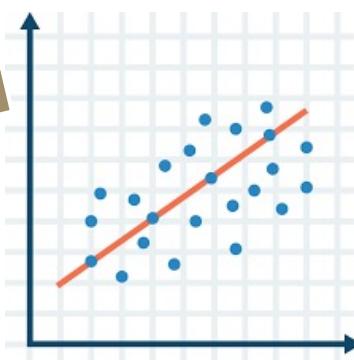
Aprendizaje  
supervisado



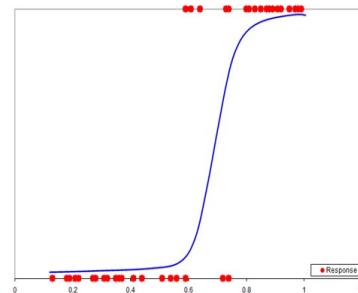
Clasificación



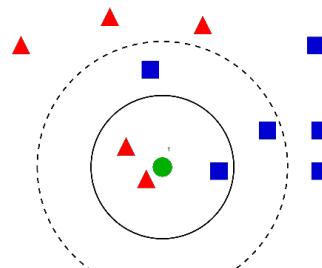
Métricas de  
Evaluación de la  
clasificación



Regresión



Regresión  
logística



KNN



Métricas de  
Evaluación de la  
regresión



# MÉTRICAS DE EVALUACIÓN

- Necesidad de evaluar la calidad de los modelos de aprendizaje automático
- Diferentes criterios a tener en cuenta:
  - **Correctitud** de la predicción
  - Simplicidad (parsimonia)
  - Interpretabilidad
  - Tiempo de aprendizaje o de predicción
  - Escalabilidad (importante para Big Data)



# MÉTRICAS DE CLASIFICACIÓN

- Una **matriz de confusión** evalua diferentes métricas de correctitud, que permiten establecer **objetivos de negocio**
- Se utilizan dos calificadores para describir cada una de sus casillas:
  - Un calificador de la correctitud de la predicción con respecto a la realidad: Verdadero o Falso
  - Un calificador del tipo de la predicción: Positivo o Negativo, con respecto a cada clase de interés (i.e churn)
- Dependiendo del contexto los tipos de error pueden ser mas costosos que otros

		Predicción	
		Churn <b>P</b>	No churn <b>N</b>
Realidad	Churn <b>+</b>	VP	FN - Tipo II
	No churn <b>-</b>	FP - Tipo I	VN

- La diagonal (en verde) muestra las instancias correctamente clasificadas. Las demás casillas resume diferentes tipos de error:
  - Tipo I: Falsos positivos
  - Tipo II: Falsos negativos

**¿Qué pasa cuando hay mas de dos clases?**



# MÉTRICAS DE CLASIFICACIÓN

- Interpretarían el caso de la detección de un email spam

TP, TN:

FP: , consecuencia:

FN: , consecuencia:

- Interpretar el caso del diagnóstico de una enfermedad grave?

TP, TN:

FP: , consecuencia:

FN: , consecuencia:

		Predicción	
		Churn P	No churn N
Realidad	Churn +	VP	FN - Tipo II
	No churn -	FP - Tipo I	VN

- Interpretar el caso de la prospección de clientes de un crédito de consumo (baja aceptación)

TP, TN:

FP: , consecuencia:

FN: , consecuencia:



# MÉTRICAS DE CLASIFICACIÓN

- Tasa de correctitud (categorization accuracy) =  $(VP+VN)/(VP+VN+FP+FN)$
- Error de mala clasificación (opposite of accuracy) =  $(FP+FN)/(VP+VN+FP+FN)$ : probabilidad de error
- Precisión=  $VP / (VP+FP)$ : valor de predicción positiva,  $P(\text{Real}+ | \text{Predicho}+)$
- Recall (o TPR o sensibilidad)=  $VP / (VP+FN)$ : qué proporción de todos los positivos reales pude identificar como tal,  $P(\text{Predicho}+ | \text{Real}+)$
- Especificidad (o TNR) =  $VN / (VN+FP)$ : qué proporción de todos los negativos reales pude identificar como tal,  $P(\text{Predicho}- | \text{Real}-)$
- Tasa de falsos positivos (o FPR) =  $FP / (VN+FP)$

		Predicción	
		Churn P	No churn N
Realidad	Churn +	VP	FN - Tipo II
	No churn -	FP - Tipo I	VN

**Medida-F (F-Score):** Promedio armónico entre precisión y recall

$$2 * (\text{Precisión} * \text{Recall}) / (\text{Precisión} + \text{Recall})$$

Imaginemos el problema de detección de spam mail e interpretemos cada métrica

Imaginemos el problema de diagnóstico de cáncer e interpretemos cada métrica

¿Cuál es la relación entre especificidad y FPR?

¿Cómo son la especificidad y sensibilidad del baseline?

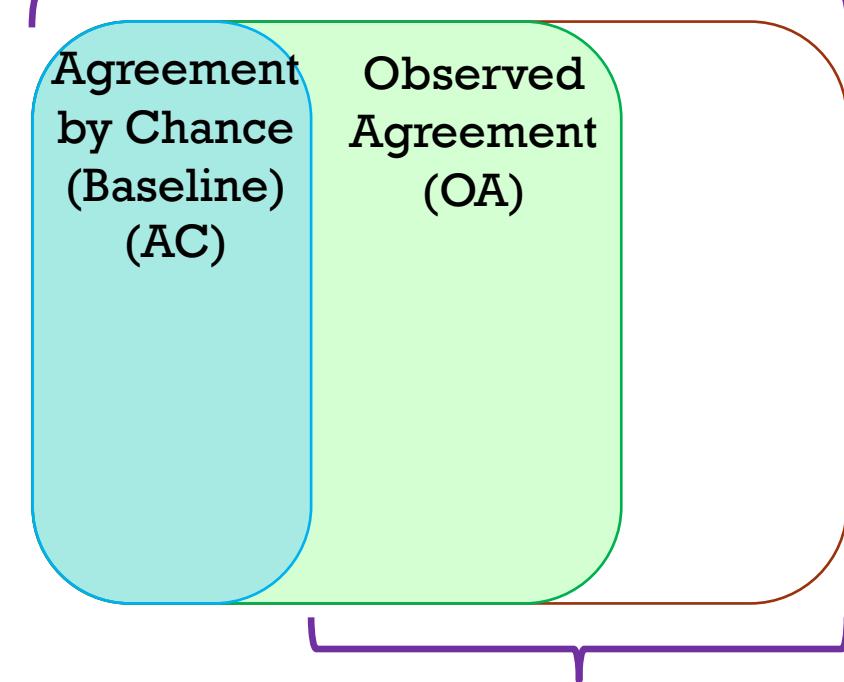


# MÉTRICAS DE CLASIFICACIÓN

## ■ Coeficiente de concordancia **Kappa**

- Para datos nominales u ordinales
- Concordancia entre las predicciones y las clases reales
- Sustrae el efecto de concordancia por suerte (AC) del valor del **accuracy** (concordancia observada - OA)
- Valores van de 0 a 1
- Muy útil sobretodo cuando las clases no están balanceadas
  - Diagnóstico de enfermedades raras
  - Clientes que acepten productos de crédito)
- $$\text{Kappa} = \frac{OA - AC}{1 - AC}$$

Modelo perfecto, accuracy=100%



Qué tanto mejor es el modelo comparado con el Baseline 1-(AC)



# MÉTRICAS DE CLASIFICACIÓN

## ■ Coeficiente de concordancia Kappa

- Para datos nominales u ordinales
- Concordancia entre las predicciones y las clases reales
- Sustrae el efecto de concordancia por suerte (AC) del valor del **accuracy** (concordancia observada - OA)
- Valores van de 0 a 1
- Muy útil sobretodo cuando las clases no están balanceadas
  - Diagnóstico de enfermedades raras
  - Clientes que acepten productos de crédito)

		Predicciones		TOTAL
		+	-	
reales	+	10	4	14
	-	3	2	5
TOTAL		13	6	19

$$\begin{aligned} \text{OA} &= 0,63 \\ \text{AC} &= 0,59 \\ \text{Kappa} &= 0,11 \end{aligned}$$

$$\begin{aligned} \text{Accuracy (OA)} &= (10+2)/19=0,63 \\ (\text{AC}) &= (13/19 * 14/19) + (6/19 * 5/19) = 0,59 \\ \text{Kappa} &= (\text{OA}-\text{AC})/(1-\text{AC}) = 0,11 \end{aligned}$$

		Predicciones		TOTAL
		+	-	
reales	+	0	3	3
	-	0	97	97
TOTAL		0	100	100

$$\begin{aligned} \text{OA} &= 0,97 \\ \text{AC} &= 0,97 \\ \text{Kappa} &= 0,00 \end{aligned}$$

$$\begin{aligned} \text{Accuracy (OA)} &= (0+97)/100=0,97 \\ (\text{AC}) &= (0/100 * 3/100) + (100/100 * 97/100) = 0,97 \\ \text{Kappa} &= (\text{OA}-\text{AC})/(1-\text{AC}) = 0 \end{aligned}$$

		Predicciones		TOTAL
		+	-	
reales	+	1475	988	2463
	-	556	1981	2537
TOTAL		2031	2969	5000

$$\begin{aligned} \text{OA} &= 0,69 \\ \text{AC} &= 0,50 \\ \text{Kappa} &= 0,38 \end{aligned}$$



# MÉTRICAS DE CLASIFICACIÓN

## TALLER: CÁLCULO DE MÉTRICAS

Los clientes son usualmente categorizados en perfiles de comportamiento de compra de productos o servicios.

Suponga que se creó un modelo para clasificar los clientes en una de 4 clases posibles (esporádico, fiel, parcial y promocional), cuya matriz de confusión presentamos a continuación:

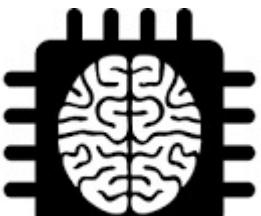
REALIDAD	PREDICCIÓN					Total
	Esporádico	Fiel	Parcial	Promocional		
Esporádico	61	8	1	0	70	
Fiel	0	56	17	0	73	
Parcial	0	0	15	0	15	
Promocional	0	0	0	24	24	
Total	61	64	33	24	182	

Calcule las métricas de evaluación de un modelo de clasificación cuyos resultados están reflejados en la tabla siguiente (ver PDF)

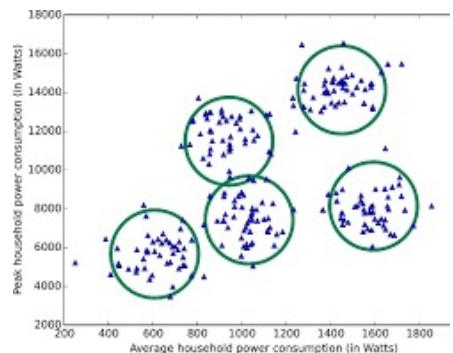


# Clase anterior

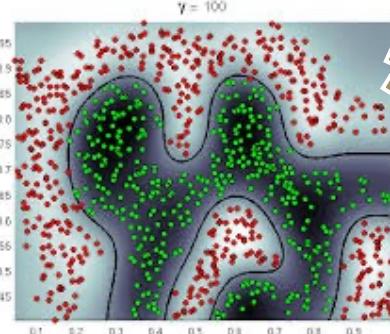
AULANDA



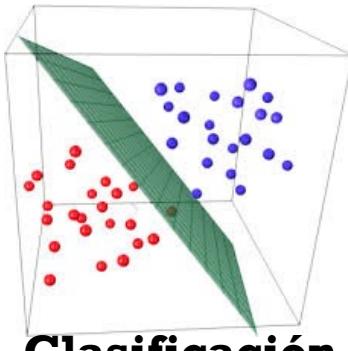
Aprendizaje  
automático



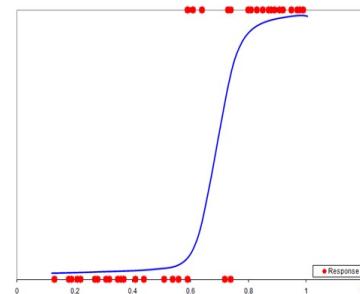
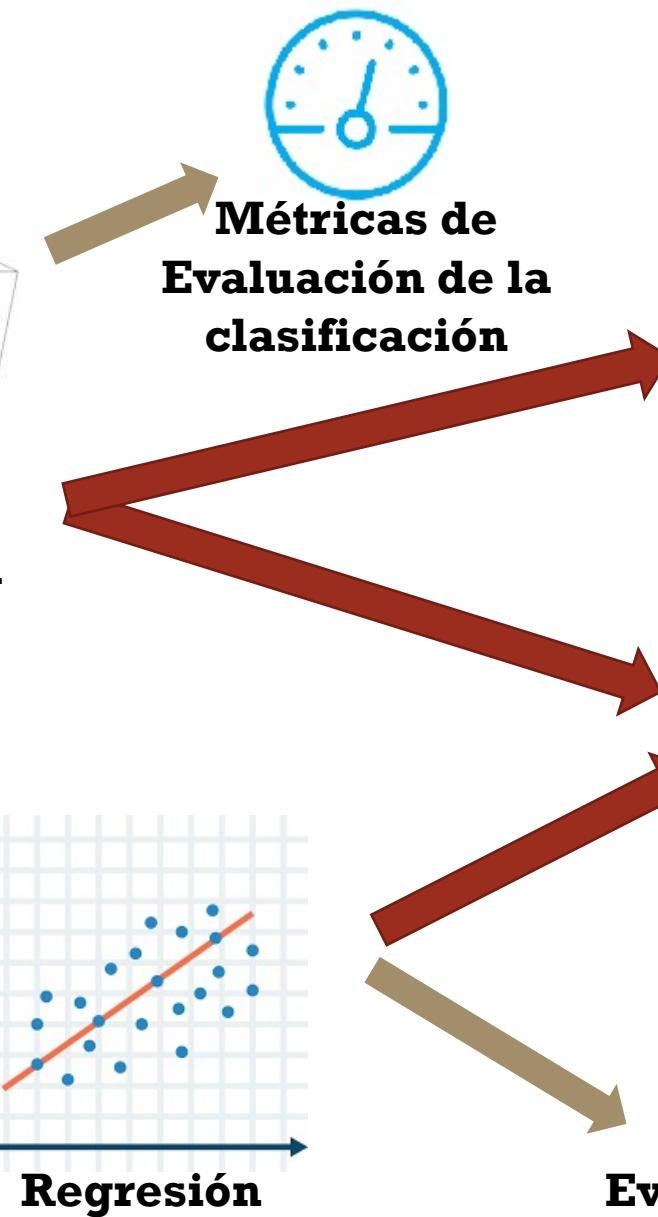
Aprendizaje  
no supervisado



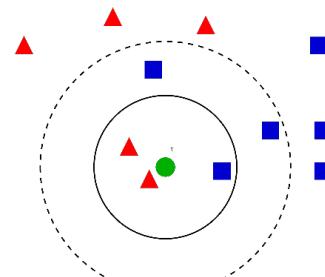
Aprendizaje  
supervisado



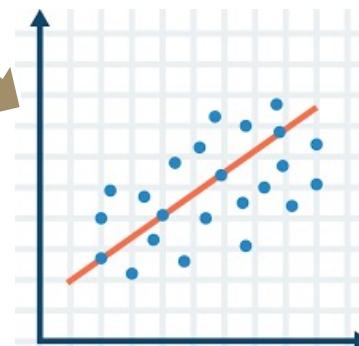
Clasificación



Regresión logística



KNN



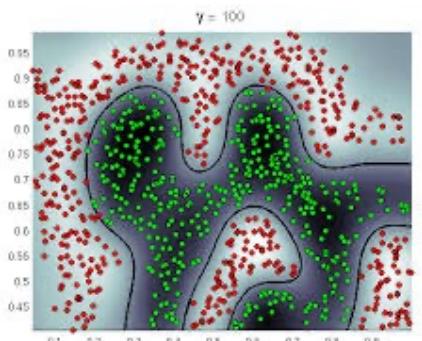
Regresión



Métricas de Evaluación de la regresión



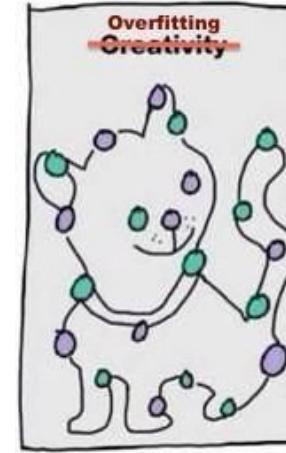
# AGENDA



**Aprendizaje  
supervisado**



**Protocolos**



**Sobre aprendizaje  
(Overfitting)**



# **OVERTFITTING Y PROTOCOLOS DE EVALUACIÓN**



# SESCO / VARIANZA

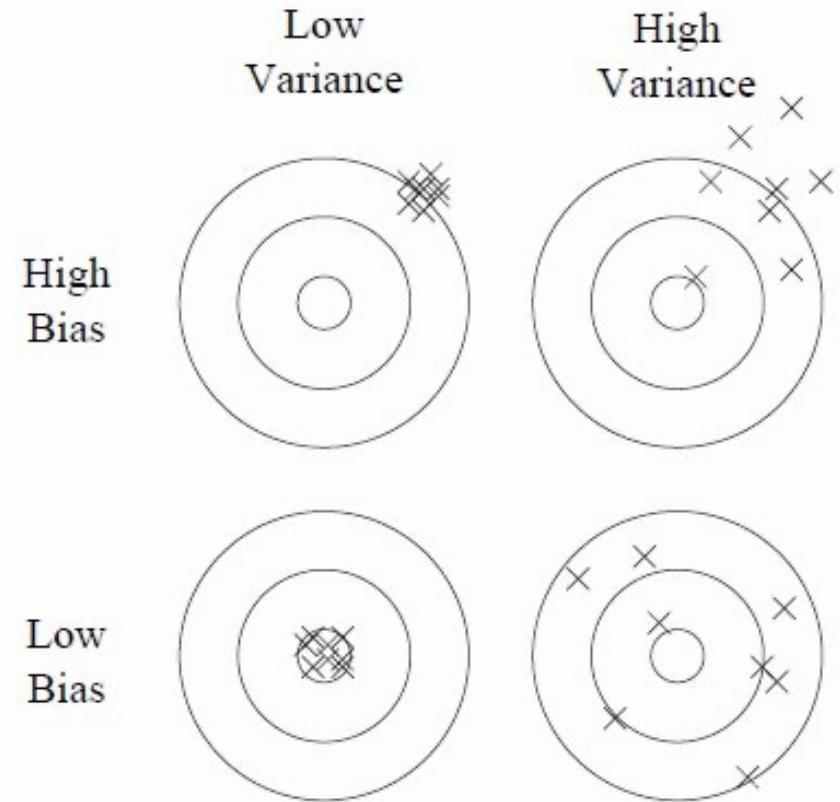
- **Sesgo** (bias): que tan lejos está el modelo de la verdad
- **Varianza**: Qué tanto varían los datos de la predicción para una misma instancia

$$Err(x) = \underbrace{\left( E[\hat{f}(x)] - f(x) \right)^2}_{\text{Sesgo}^2} + \underbrace{E\left[ \hat{f}(x) - E[\hat{f}(x)] \right]^2}_{\text{Varianza}} + \sigma_e^2$$

cada resultado del modelo

Promedio del modelo

Error irreducible

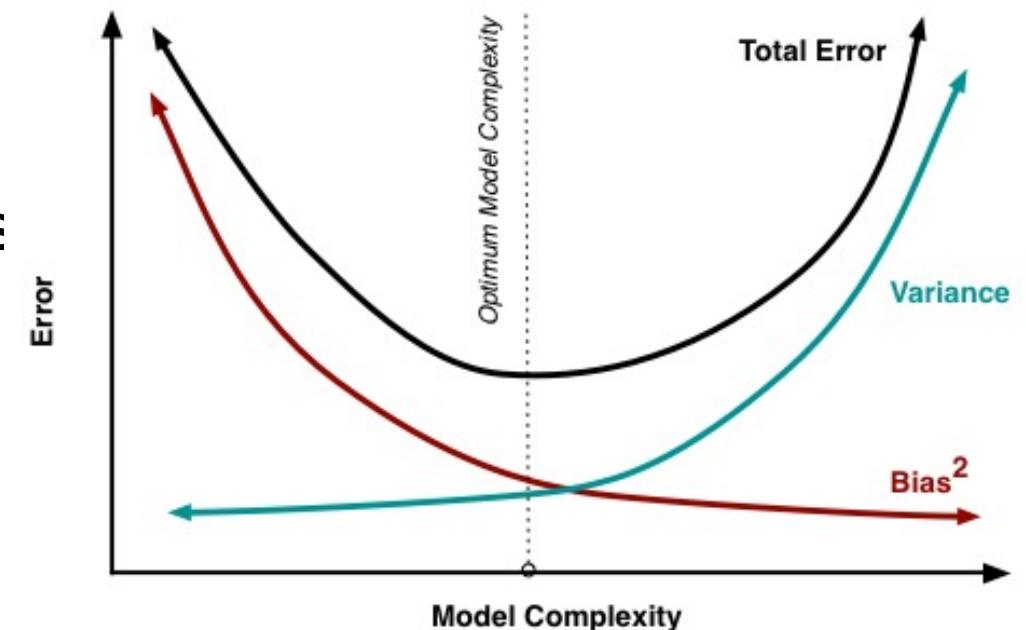


Domingo, 2012



# SESCO / VARIANZA

- Ambos son fuente de error
- Se debe determinar un **compromiso** entre ambos tipos de error
- Parámetros de los modelos controlan la complejidad



<http://scott.fortmann-roe.com/docs/BiasVariance.html>



# SOBRE APRENDIZAJE (OVERFITTING)

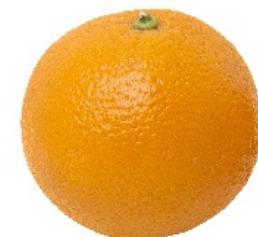
¿Cómo le enseño a mi hijo que es una pelota?

Set de entrenamiento



¿Qué patrones distinguen las pelotas de los demás juguetes?

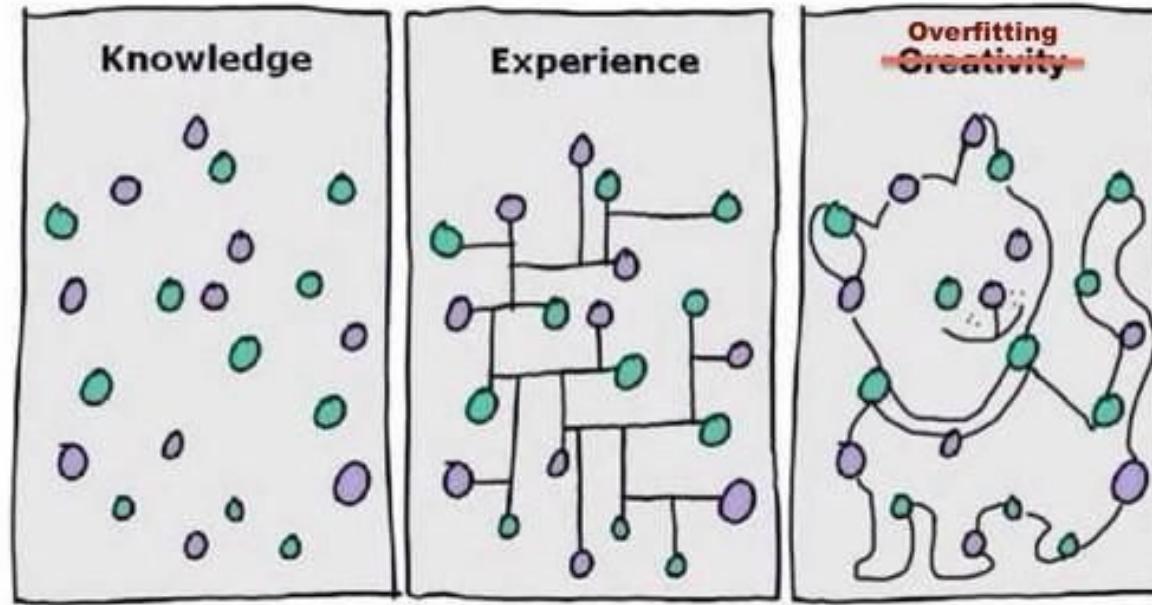
Hijo, ¿Es esta una pelota?



¿Cómo caracterizo una situación de un modelo con overfitting?



# SOBRE APRENDIZAJE (OVERFITTING)



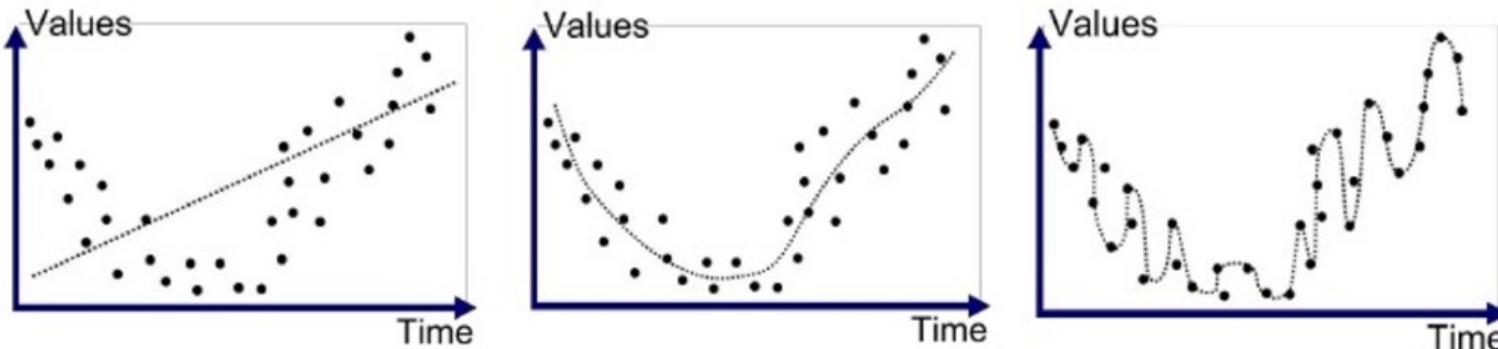
<http://blog.algotrading101.com/design-theories/what-is-curve-fitting-overfitting-in-trading/>

- **Sobre aprendizaje:** Los modelos aprenden a describir los errores aleatorios o el “ruido” del conjunto de entrenamiento.
- Ocurre cuando un modelo se vuelve excesivamente **complejo**

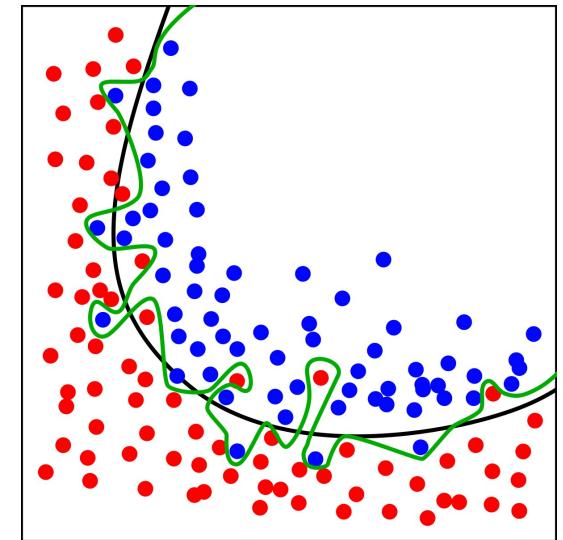


# SOBRE APRENDIZAJE (OVERFITTING)

Regresión



Clasificación



¿Cómo es el sesgo y la varianza de estos modelos?

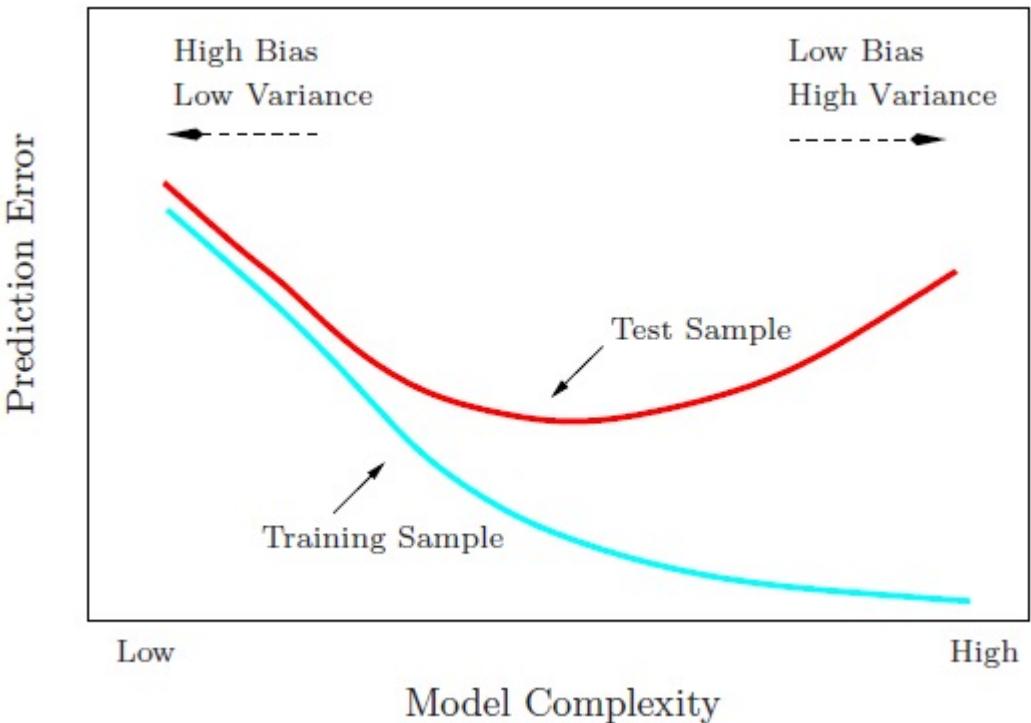
- La **complejidad** de un modelo debe ajustarse de tal manera que permita la **generalización**, al utilizarse con datos que no haya conocido durante el proceso de entrenamiento
- Principio de **parsimonia** (Occam's Razor): la mejor explicación es la mas simple → preferir los modelos más sencillos con menos suposiciones

<https://en.wikipedia.org/wiki/Overfitting>



# SOBRE APRENDIZAJE (OVERFITTING)

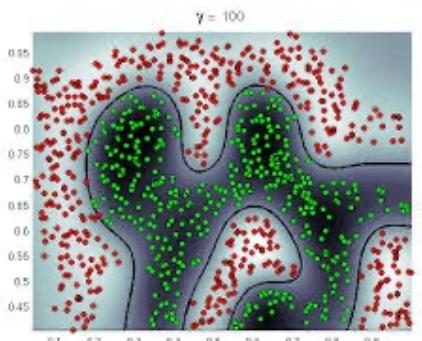
- Los modelos tienden a ajustarse al conjunto de datos usado para su aprendizaje → el **error de entrenamiento** es un mal estimador
- Queremos encontrar la complejidad del modelo que nos permita minimizar el **error de test**



<https://onlinecourses.science.psu.edu/stat857/node/160>



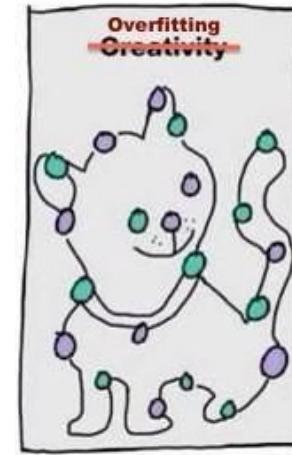
# AGENDA



**Aprendizaje  
supervisado**



**Protocolos**



**Sobre aprendizaje  
(Overfitting)**



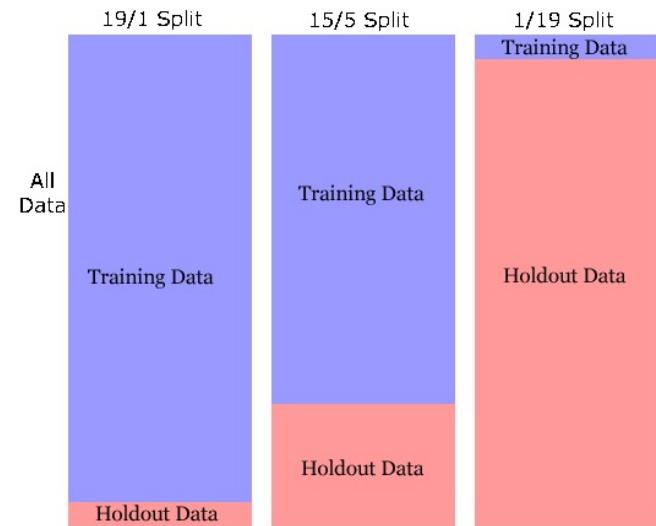
# PROTOCOLOS DE EVALUACIÓN

- Aplican para aprendizaje supervisado en general (tanto para clasificación como para regresión).
- Permiten definir los mejores valores de los parámetros de los modelos en términos de **flexibilidad**.
- Evaluar cual sería la capacidad de **generalización** del modelo a datos nuevos
- Diferenciar entre el **error de entrenamiento** y el **error de test**. Evitar el sesgo causado por la **subestimación del error** al evaluar con el mismo set de entrenamiento.
- Permitir establecer un compromiso entre sesgo y varianza, luchando contra el **sobre aprendizaje**, en busca de un modelo con buenas **capacidades predictivas**

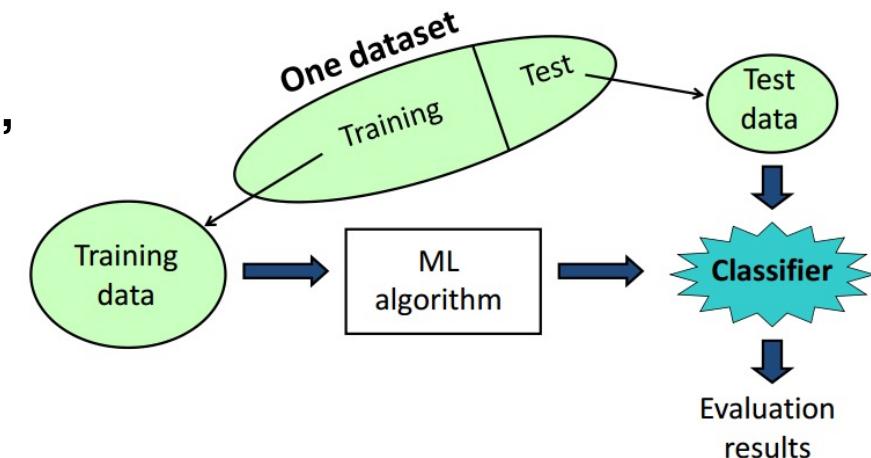


# PROTOCOLOS DE EVALUACIÓN

- **Holdout:** particionar el conjunto de datos en 2:
  - **Conjunto de entrenamiento:** con el que se aprende el algoritmo de clasificación
  - **Conjunto de validación o test:** separa al comienzo del procedimiento y no se considera en el aprendizaje
  - **Aleatoriedad** del particionamiento
  - **Compromiso:** entre mas datos mejor el aprendizaje, entre mas datos mejor la evaluación
- **Repeated holdout:** repetir el procedimiento y agregar las métricas de evaluación



<https://webdocs.cs.ualberta.ca/~aixplore/learning/DecisionTrees/InterArticle/6-DecisionTree.html>



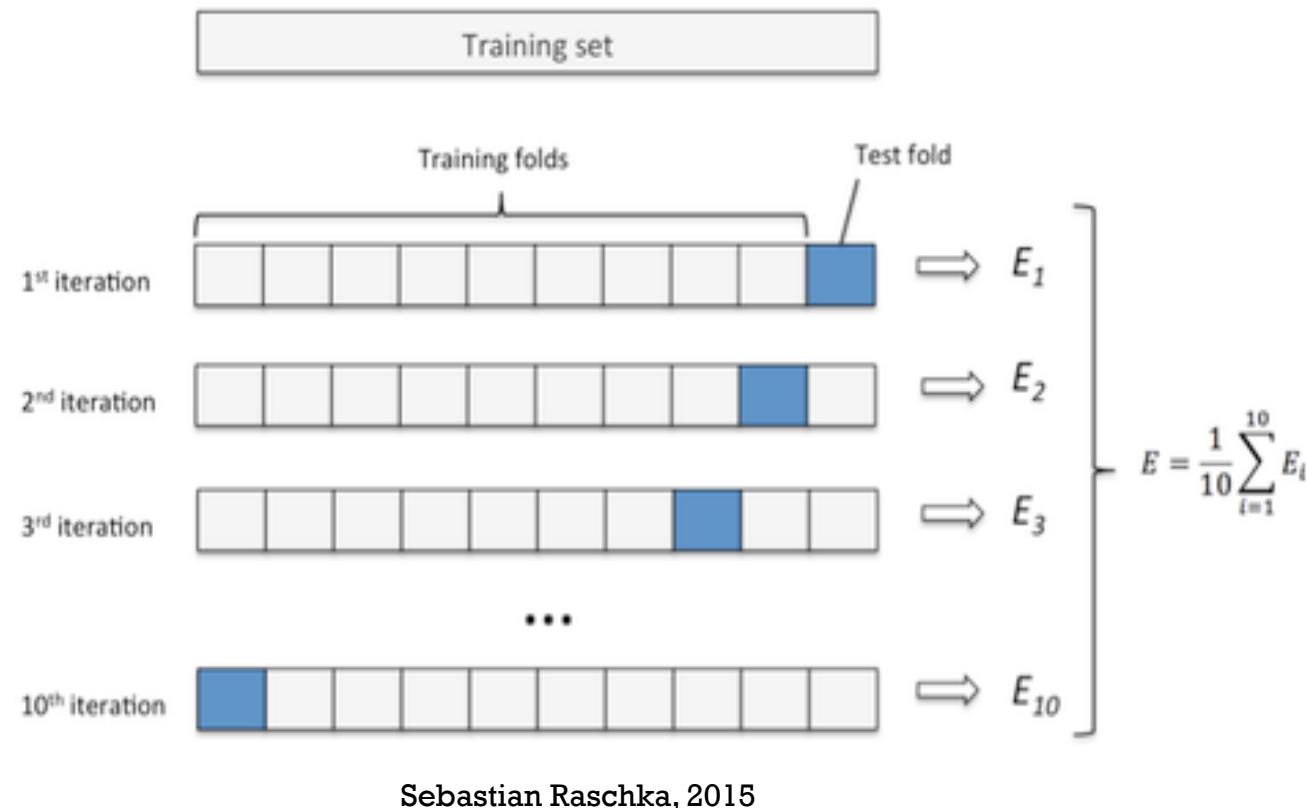
Ian Witten, Weka MOOC



# PROTOCOLOS DE EVALUACIÓN

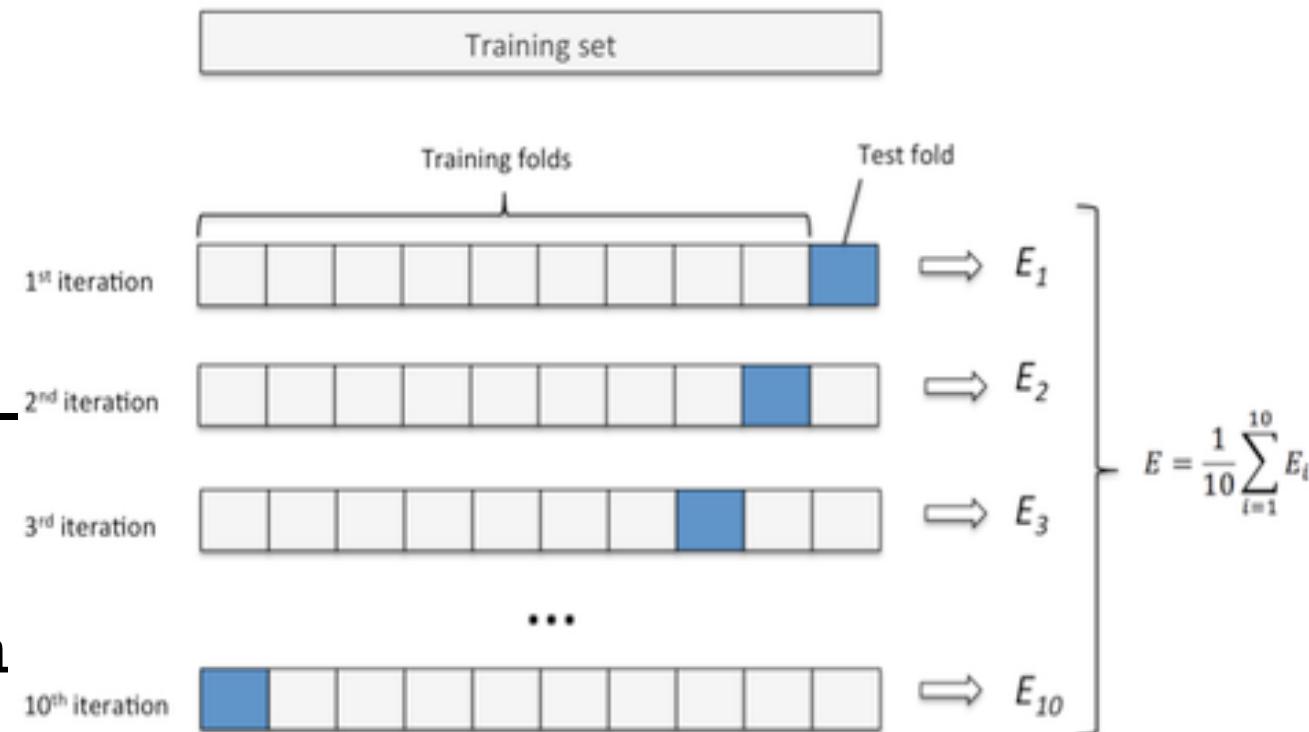
## ■ **K-fold cross-validation:**

- Particionar el set de datos en **K** conjuntos disyuntos del mismo tamaño
- **K-1** partes se usan para entrenamiento, 1 parte se usa para el test
- Se repite el proceso **K** veces
- Se agregan las métricas de evaluación



# PROTOCOLOS DE EVALUACIÓN

- **K-fold cross-validation,**  
**Escogencia del K:**
  - Permite balancear entre sesgo y varianza
  - **LOOCV** (Leave One Out Cross-Validation): partes de tamaño 1
  - Por defecto se estima que los mejores resultados se obtienen con un valor de K entre 5 y 10



Sebastian Raschka, 2015



# PROTOCOLOS DE EVALUACIÓN

## ▪ **Bootstrapping:**

- Consideración de varios conjuntos de entrenamiento/test utilizando muestreo con remplazo
- Por lo general muestreros del mismo tamaño del conjunto original
- Es un muy buen estimador de los parámetros, pero no de las métricas de calidad de los modelos dado un sesgo causado por el número de observaciones distintas promedio ( $0.632^N$ )

Original Dataset

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$
-------	-------	-------	-------	-------	-------	-------	-------	-------	----------

Bootstrap 1

$x_8$	$x_6$	$x_2$	$x_9$	$x_5$	$x_8$	$x_1$	$x_4$	$x_8$	$x_2$
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Bootstrap 2

$x_{10}$	$x_1$	$x_3$	$x_5$	$x_1$	$x_7$	$x_4$	$x_2$	$x_1$	$x_8$
----------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Bootstrap 3

$x_6$	$x_5$	$x_4$	$x_1$	$x_2$	$x_4$	$x_2$	$x_6$	$x_9$	$x_2$
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

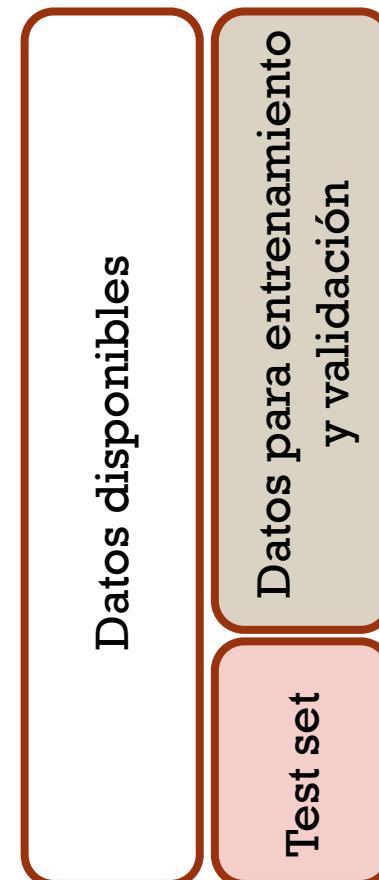
Sebastian Raschka, 2015



# PROTOCOLOS DE EVALUACIÓN

## ▪ Set de validación vs set de test:

- Separación de un set de datos de **test** para evaluación final del modelo escogido
- **Overfitting** si se calibran los modelos con el mismo set de test



- Calibración de pretratamientos (normalización, imputación)
- Calibración de los parámetros de los modelos (KFCV, holdout)
- Comparación de los resultados de diferentes modelos
- Evaluar la capacidad de generalización



# QUIZ 2



# EJEMPLO DE KNN

- 04-KNN-Protocolos-Ejemplo
- Desarrollo de k-nn sobre el dataset de iris utilizando diversos protocolos de evaluación



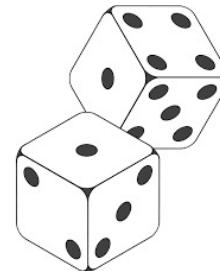
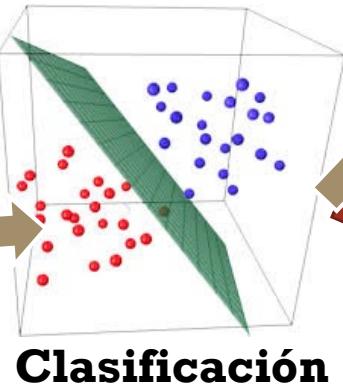
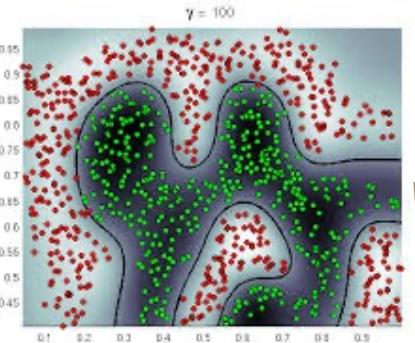
# TALLER DE EVALUACIÓN DE UN MODELO DE CLASIFICACIÓN (KNN)

- DATASET: base de datos de 20000 clientes que han cancelado (churn) o no los servicios de una compañía. La idea es poder predecir en un futuro quiénes son los clientes más propensos a hacer churn, para poder desarrollar campañas que lo prevengan.
- Descarguen los archivos del taller de clasificación de churn de clientes y ejecuten cada una de las partes.
- Documento 04-KNN-CHURN.html



# **APRENDIZAJE SUPERVISADO**





## Probabilidades

THE PROBABILITY OF "B" BEING TRUE GIVEN THAT "A" IS TRUE

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

↑  
THE PROBABILITY  
OF "A" BEING TRUE  
GIVEN THAT "B" IS  
TRUE

↓  
THE PROBABILITY  
OF "A" BEING  
TRUE

↖  
THE PROBABILITY  
OF "B" BEING  
TRUE

## Naïve Bayes



# PRINCIPIOS BASICOS DE PROBABILIDAD

Probabilidad Marginal:

$$P(X)$$

Regla del producto: *(Probabilidad Conjunta)*

$$P(X \cap Y) = P(X) * P(Y) \text{ (si } X \text{ y } Y \text{ son eventos independientes)}$$

Probabilidad condicionada:

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$$

Ley de probabilidad total:

$$P(X) = \sum_n P(X|Y_n) P(Y_n) \text{ (si } X \text{ y } Y_n \text{ son eventos independientes)}$$

Regla de Bayes:

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} = \frac{P(Y|X) * P(X)}{P(Y)}$$

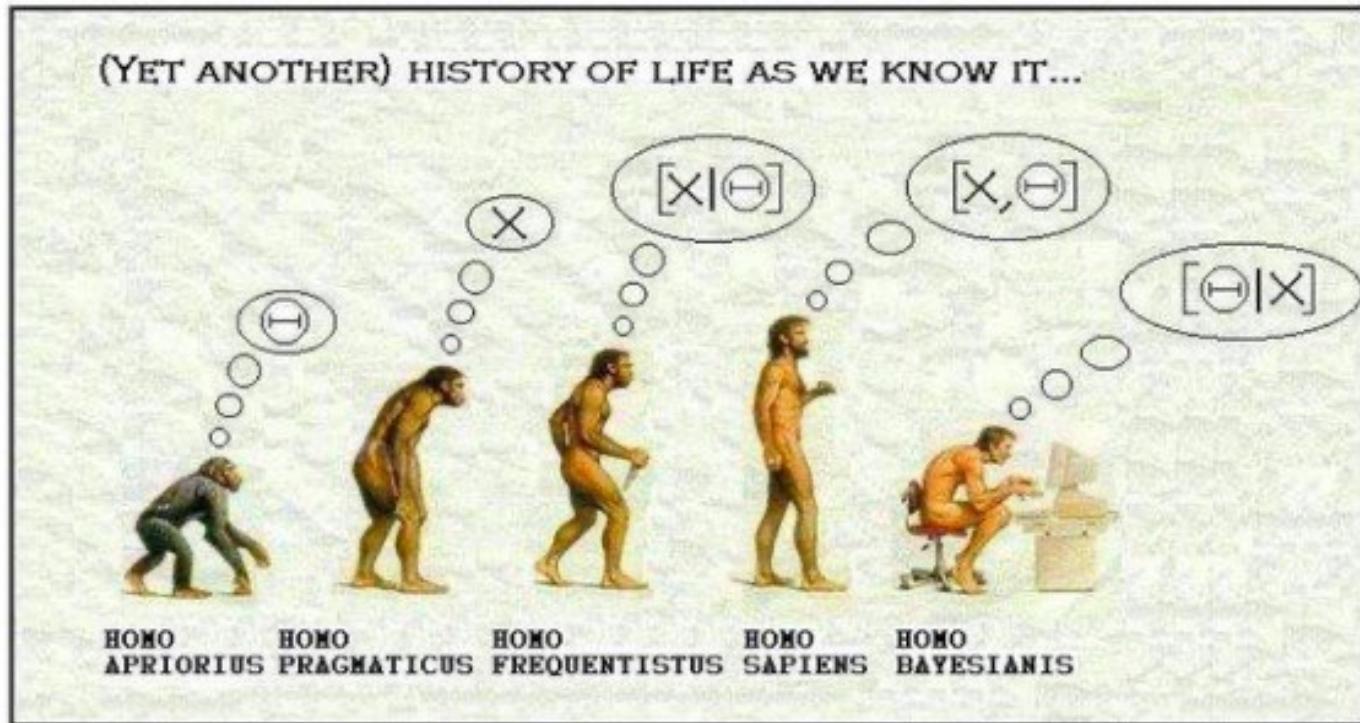


# NAIVE BAYES: TALLER

Descarguen el taller de Excel y PDF de Naive Bayes que se encuentra en la sesión 5

Desarrollen las partes 1 y 2 del taller de repaso, correspondientes al cálculo de probabilidades básicas y probabilidad condicional







La estimación de la probabilidad de un evento, o un resultado potencial, debe basarse en la evidencia dada por múltiples ensayos u oportunidades para que ocurra el evento



Los métodos bayesianos proporcionan información sobre cómo la probabilidad de estos eventos puede ser estimada a partir de los datos observados



Los principios básicos de probabilidad se usan transversalmente en el algoritmo **Naïve Bayes**

# PROBABILIDAD BAYESIANA



# CLASIFICADORES BAYESIANOS

- Los **clasificadores bayesianos** asignan cada observación a la clase  $j$  más probable, dados los valores observados de sus variables predictivas:  
$$\text{argmax}_j p(Y = y_j | X = x_{\text{observados}})$$
- Si se conocen las distribuciones de probabilidad, el clasificador resultante da la frontera de separación óptima en términos de error
- No siempre se tienen las probabilidades condicionales necesarias.
- **Naïve Bayes** es un algoritmo basado en el Teorema de Bayes

- **Algunas aplicaciones de clasificadores Bayesianos son:**
  - **Clasificación de texto**, como el filtrado de correo no deseado (spam)
  - Detección de intrusiones o anomalías en redes informáticas.
  - Diagnóstico de afecciones médicas debido a un conjunto de síntomas observados.
- Funcionan muy bien en problemas en los que la información de numerosos atributos deben considerarse simultáneamente para estimar la probabilidad general de un resultado

ISLR, 2013



# NAIVE BAYES (BAYES INGENUO)

Teorema de Bayes:

$$p(y_j|x_1, x_2, \dots, x_n) = \frac{p(y_j, x_1, x_2, \dots, x_n)}{p(x_1, x_2, \dots, x_n)} = \frac{p(y_j) * p(x_1, x_2, \dots, x_n | y_j)}{p(x_1, x_2, \dots, x_n)}$$

- El denominador es solo usado para propósitos de normalización (suma de probabilidades = 1)

$$p(x_1, x_2, \dots, x_n) = \sum_j p(y_j) * p(x_1, x_2, \dots, x_n | y_j)$$

- Por ello solo nos fijamos en el numerador:

$$p(y_j, x_1, x_2, \dots, x_n) = p(y_j) * p(x_1|y_j) * p(x_2|x_1, y_j) * p(x_3|x_2, x_1, y_j) * \dots * p(x_n|x_{1:n-1}, y_j)$$

- Si asumimos ingenuamente (**naïvely**) que todas las variables predictivas  $x_i$  son independientes condicionalmente con respecto a la clase  $y_j$ <sup>1</sup> entonces el numerador se simplifica a:

$$\begin{aligned} & p(y_j) * p(x_1|y_j) * p(x_2|y_j) * p(x_3|y_j) * \dots * p(x_n|y_j) \\ &= p(y_j) \prod_{i=1}^n p(x_i|y_j) \end{aligned}$$



# NAÏVE BAYES (BAYES INGENUO)

- La regla de clasificación es:

$$\operatorname{argmax}_j p(y_j) \prod_{i=1}^n p(x_i|y_j)$$

- **Sólo necesitamos especificar :**

→ Las probabilidades a priori de cada clase

→ Las distribuciones de probabilidad de las variables predictivas para cada clase (condicionadas a la clase)

- Esta información se constituye en los **parámetros** del modelo, y en el caso de variables categóricas se obtienen a partir de tablas de frecuencias (conteos)



# NAIVE BAYES

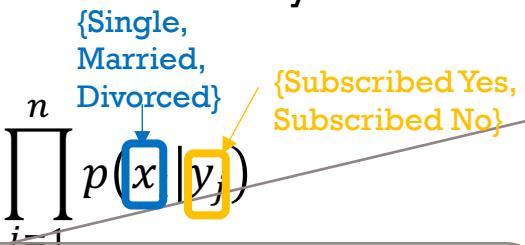
**Ejemplo:** Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Naïve Baye a partir de los datos históricos para calcular las probabilidades posteriores de cada clase: `subscribed=yes` and `subscribed=no`.

$$p(y_j|x_1, \dots, x_n) = \operatorname{argmax}_j p(y_j)$$

Marital	Subscribed=yes
Single	35%
Married	53%
Divorced	12%

Subscribed=yes      12%



Marital	Subscribed=no
Single	28%
Married	61%
Divorced	11%

Subscribed=no      88%

¿Debería el banco ofrecerle un CDT al cliente con la información siguiente?

Job=Management  
Marital=Married  
Education=Secondary  
Default=no  
Housing=yes  
Loan=no  
Contact=Cellular  
Outcome=Success

Suponga que se disponen de las probabilidades condicionales para todas las variables predictivas (ya ilustradas para el estado civil "Marital")



# NAIVE BAYES

**Ejemplo:** Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Naïve Bayes a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: subscribed=yes and subscribed=no.

$$p(y_j|x_1, \dots, x_n) = \operatorname{argmax}_j p(y_j) \prod_{i=1}^n p(x_i|y_j)$$

Marital	Subscribed=yes	Marital	Subscribed=no
Single	35%	Single	28%
Married	53%	Married	61%
Divorced	12%	Divorced	11%
Subscribed=yes		Subscribed=no	
12%		88%	

¿Debería el banco ofrecerle un CDT al cliente con la información siguiente?

	Subscribed=yes	Subscribed=no
Job=Management	22%	21%
Marital=Married	53%	61%
Education=Secondary	46%	51%
Default=no	99%	98%
Housing=yes	35%	57%
Loan=no	90%	85%
Contact=Cellular	85%	62%
Outcome=Success	15%	1%
Priors	12%	88%

Numerador	0,000255914	0,000169244
Proba posterior	60%	40%



# NAÏVE BAYES (BAYES INGENUO)

- ¿Qué pasa si algunos de los valores de las variables predictivas tienen frecuencia nula con respecto a las categorías de la clase? ¿cuáles serían sus probabilidades a posteriori asociadas?
- Para evitar este problema, se utilizan métodos de **suavización**.
  - Por ejemplo, al contar las frecuencias de ocurrencia de cada valor se les agrega un valor pequeño,  $\varepsilon > 0$ , evitando que alguna probabilidad sea cero:

$$P(\text{casado}|\text{cliente potencial}) = \frac{\text{Conteo}(\text{casado}, \text{cliente potencial}) + \varepsilon}{\text{Conteo}(\text{cliente potencial}) + N(x) * \varepsilon}$$

- El método de suavización de **Laplace** se aplica usualmente con  $\varepsilon=1$ , otro valor puede ser  $1/n$  donde  $n$  es el número de datos de entrenamiento.



# NAÏVE BAYES (BAYES INGENUO)

- Cuando las variables predictivas no son categóricas (e.g. numéricas), es necesario establecer una distribución de probabilidad:
  1. Se puede discretizar (en compartimentos) la variable convirtiéndola en categórica.
  2. Se establece una distribución de probabilidad empírica utilizando KNN,

$$P(Y = j|X = x_0) = \frac{1}{k} \sum_{i \in \mathcal{N}_0} I(y_i = j)$$

- 3. Se supone que se trata de un tipo de distribución de probabilidad y se utiliza su función de densidad.
- Por ejemplo, si se supone la variable sigue una **distribución normal** condicionada a la categoría objetivo, se puede calcular la media  $\mu$  y desviación estándar  $\sigma$  a partir de los datos históricos, y utilizar la función de densidad:

$$P(edad|cliente\ potencial) = \frac{1}{\sigma_{edad|cliente}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{edad - \mu_{edad|cliente}}{\sigma_{edad|cliente}}\right)^2}$$



# NAÏVE BAYES (BAYES INGENUO)

## Pros:

- **Simple, rápido** y muy **efectivo**, permite atributos tanto categóricos como numéricos
- Estima efectivamente **las probabilidades condicionales** con respecto a los valores de la categoría objetivo
- Trabaja bien con atributos categóricos con **valores faltantes** y con ruido
- Resistente al **overfitting**, sobretodo si se incluye un suavizador (e.g. Laplace)
- Trabaja bien con pequeñas muestras de entrenamiento, pero también con grandes

## Contras:

- Sólo se puede utilizar para **clasificación**
- Se basa en **suposiciones** muy fuertes (igualdad de importancia de las características independientes )
- **No ideal para datasets con muchos atributos numéricos**
- **Muy sensible** a atributos correlacionados (considera varias veces los mismos efectos)
- Las probabilidades estimadas son menos confiables que las clases predichas



# NAIVE BAYES: TALLER

## 1. Continuando con el taller de Naive Bayes:

Desarrollen la parte 3, aplicación de Bayes ingenuo para dos variables predictivas categóricas, y

La parte 4, de aplicación de Bayes ingenuo para variables predictivas numéricas



# EJEMPLO DE NAIVE BAYES

- 05-Ejemplo-Bayes
  - Desarrollo del naïve bayes desde cero
  - Naive\_bayes de sklearn



# TALLER: NAIVE BAYES, APLICACIÓN

- Desarrollar del taller de Naive Bayes, que se encuentra especificado en el documento 05-01-NaiveBayes-Iris.html



# REFERENCIAS

- *Introduction to Statistical Learning with Applications in R (ISLR)*, G. James, D. Witten, T. Hastie & R. Tibshirani, 2014
- *Machine Learning with R*, Brett Lantz, Packt Publishing, 2015
- *Machine Learning*, Tom M. Mitchell, McGraw-Hill, 1997

