

UNIVERSIDADE DA CORUÑA

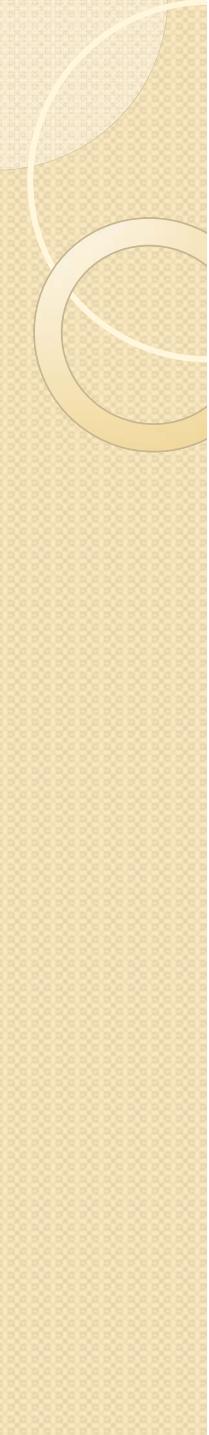
APRENDIZAJE AUTOMÁTICO

TEMA 5:
OTROS CONCEPTOS



EVALUACIÓN

- Comparar los resultados de varios algoritmos
 - De dos algoritmos distintos
 - Para evaluar cuál se comporta mejor en un determinado problema
 - Del mismo algoritmo con distintos parámetros
 - Para evaluar qué configuración de parámetros es la más indicada
- Si los algoritmos son determinísticos se puede comparar el nivel de error / precisión / etc.



EVALUACIÓN

- Si no son determinísticos, no se puede tomar el resultado de una única ejecución
 - Existirán variaciones en el resultado debido precisamente a esta componente aleatoria
 - Para minimizar la componente aleatoria, es necesario repetir el experimento un número alto de veces (50-100 veces) y quedarse con la media de los resultados
 - Esa media se usará como base para comparar los resultados



EVALUACIÓN

- Incluso si son determinísticos, no se puede tomar el resultado de una única ejecución...
 - Resultado en el conjunto de test
- **...si el conjunto de test ha sido escogido de forma aleatoria**
 - Si el conjunto de test es siempre el mismo y los algoritmos son determinísticos, el resultado en test será siempre el mismo
 - Ejecutarlo una única vez



EVALUACIÓN

- Incluso si son determinísticos, no se puede tomar el resultado de una única ejecución...
 - Resultado en el conjunto de test
- **...si el conjunto de test ha sido escogido de forma aleatoria**
 - Se introduce una componente aleatoria que hace que el proceso global no sea determinístico
 - Al igual que antes, para minimizar la componente aleatoria, es necesario repetir el experimento un número alto de veces (50-100 veces) y quedarse con la media de los resultados
 - Esa media se usará como base para comparar los resultados



EVALUACIÓN

- Por tanto, se tienen 2 fuentes de variabilidad:
 - Elección del conjunto de entrenamiento y test
 - Variabilidad propia del algoritmo
 - Por ejemplo, inicialización aleatoria de pesos en una RNA
 - Hay algoritmos que la tienen, y algoritmos que no
- Necesario minimizar ambas fuentes de variabilidad
 - Repetir el experimento para minimizar la componente aleatoria
 - Quedarse con la media de los resultados
 - Esa media se usará como base para comparar los resultados



EVALUACIÓN

- Cómo minimizar ambas fuentes de variabilidad (1/2)
 - Si los conjuntos de entrenamiento/test son fijos
 - Se especifican de forma separada los conjuntos de entrenamiento y test
 - Si el sistema es determinístico, ejecutarlo una vez
 - Si el sistema no es determinístico, ejecutarlo varias veces
 - Promediar los resultados
 - Como resultado, se tendrá un valor para estos conjuntos de entrenamiento/test fijos

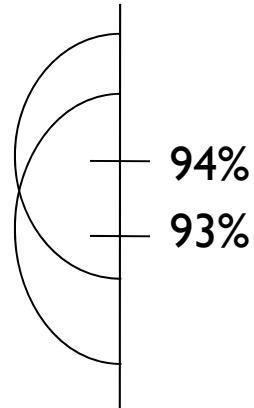


EVALUACIÓN

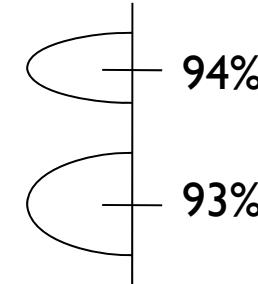
- Cómo minimizar ambas fuentes de variabilidad (2/2)
 - Si los conjuntos de entrenamiento/test se escogen de forma aleatoria
 - Se parte de un conjunto de patrones sin especificar cuáles estarán en entrenamiento y cuáles estarán en test
 - Dividir este conjunto en entrenamiento/test varias veces
 - Para cada división
 - Si el sistema es determinístico, ejecutarlo una vez
 - Si el sistema no es determinístico, ejecutarlo un número alto de veces
 - Promediar los resultados de estas ejecuciones para este conjunto de entrenamiento
 - Como resultado, se tendrá un único valor para esta división entrenamiento/test
 - Promediar los resultados de todas las divisiones

EVALUACIÓN

- Sin embargo, no sólo basta con la media
 - Es posible que dos algoritmos den una precisión de medias 93% y 94%
 - Supondríamos que el segundo funciona mejor
 - Pero es posible que la desviación típica sea...



Un test estadístico podría decir
que los resultados son
estadísticamente similares



Un test estadístico diría que los
resultados son estadísticamente
distintos



EVALUACIÓN

- Es necesario usar la media y la desviación típica obtenida de las múltiples ejecuciones de cada algoritmo para realizar un contraste de hipótesis
 - Media y desviación típica del resultado obtenido en el conjunto de test!!!
 - Partición del conjunto de patrones en entrenamiento y test
 - Se introduce un componente aleatorio que hay que minimizar
 - Repetir el experimento



EVALUACIÓN

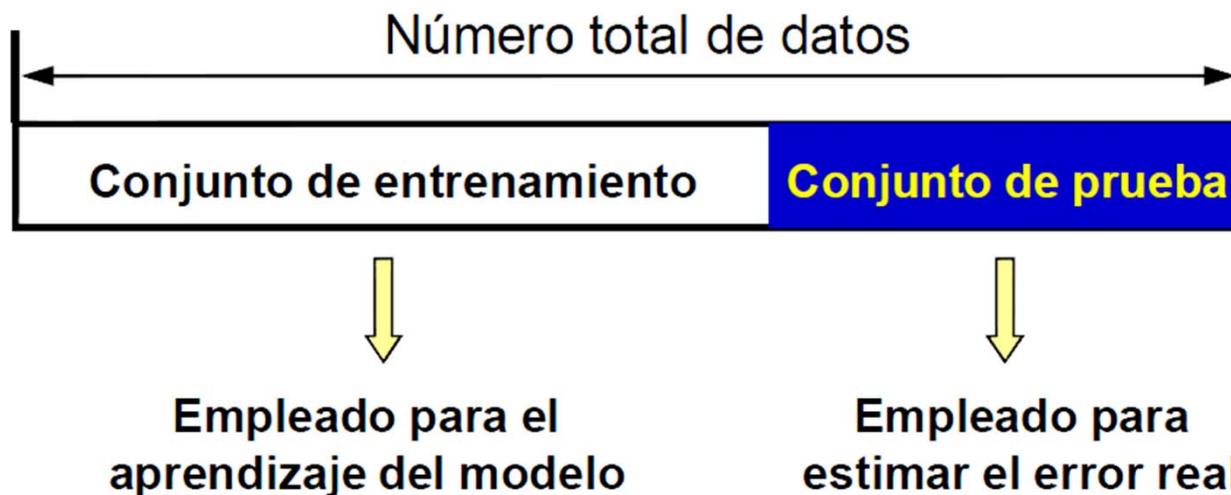
- ¿Cuántos datos escoger para el conjunto de test?
 - Importante que sea **representativo**
 - Si se tienen pocos patrones (~1000, 5000, etc), se puede tomar un 15%-20% para el conjunto de test (~200, 1000, etc.)
 - Entrenamiento/validación/test: 60/20/20%, 70/15/15% o similar
 - Si se tienen muchos patrones (100.000, 1M, etc.) (Big Data, Deep Learning), se suele tomar un porcentaje mucho menor (1%, 0.5%)
 - Entrenamiento/validación/test: 98/1/1%, 99/0.6/0.4% o similar
 - El número de patrones resultante de test es parecido al anterior
 - Más importante que sean muchos es que sean **representativos**
- ???Cómo escoger el conjunto de test???
 - Varias técnicas disponibles

EVALUACIÓN

- Técnicas:

- *Hold Out:*

- Realiza un experimento realizando una única partición
 - Del conjunto de datos disponible se selecciona aleatoriamente dos conjuntos mutuamente excluyentes (entrenamiento y prueba)
 - Partición simple del conjunto de datos

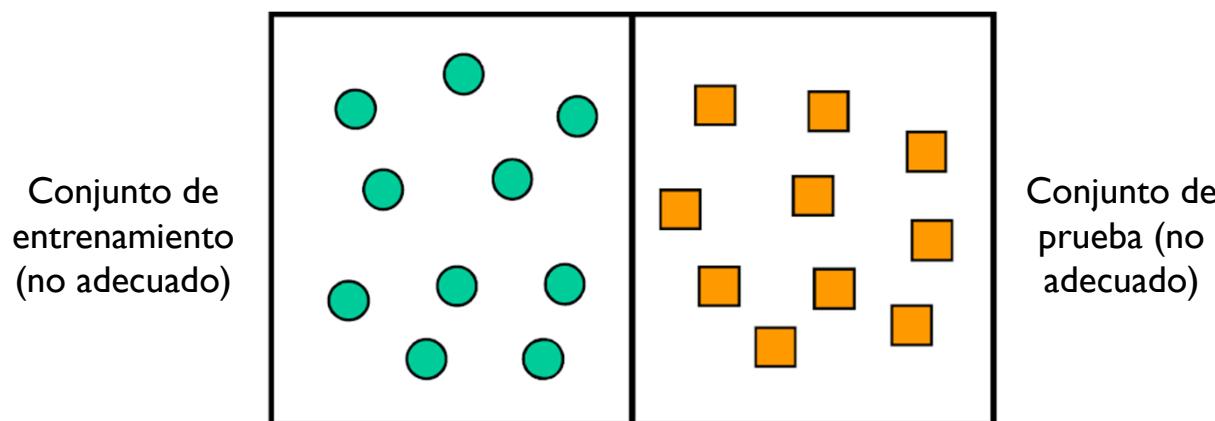


EVALUACIÓN

- Técnicas:

- *Hold Out*:

- Es considerado como una evaluación pesimista
 - El número de datos reservados para entrenar es pequeño comparado con el total
 - Puesto que sólo se realiza un único experimento con un conjunto de entrenamiento, el resultado es **engañoso** si la partición no es adecuada





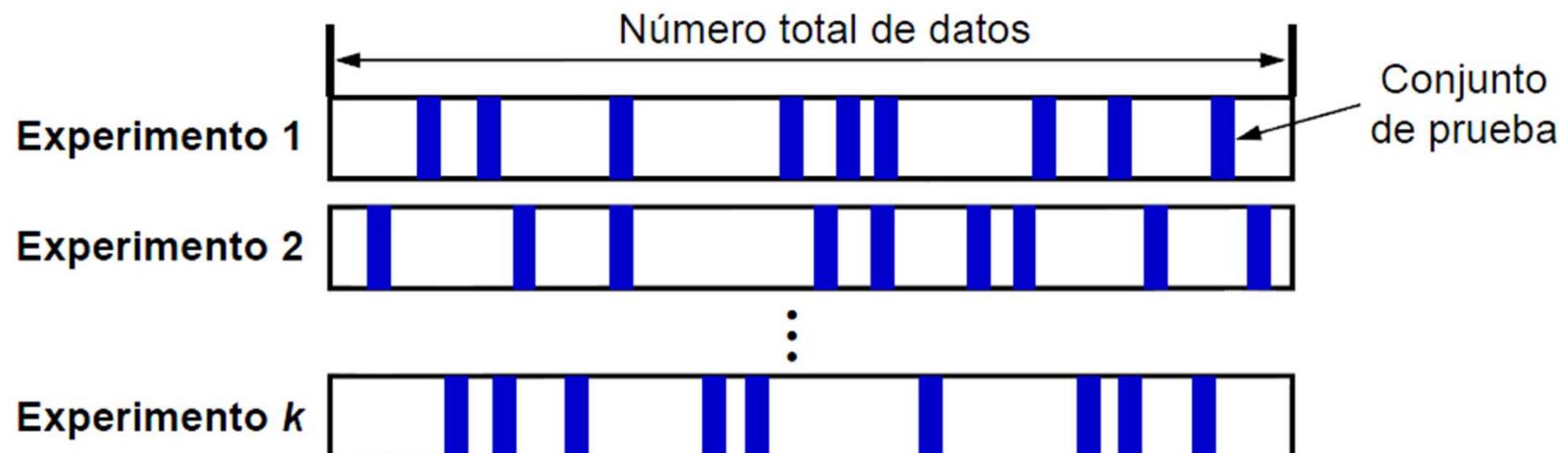
EVALUACIÓN

- Técnicas:
 - *Hold Out:*
 - Es considerado como una evaluación pesimista
 - El número de datos reservados para entrenar es pequeño comparado con el total
 - Puesto que sólo se realiza un único experimento con un conjunto de entrenamiento, el resultado es **engañoso** si la partición no es adecuada
 - Los resultados son muy dependientes de la aleatoriedad de la partición en entrenamiento y test
 - Esta fuente de variabilidad no se ha minimizado
 - Repetir el experimento (submuestreo aleatorio).

EVALUACIÓN

- Técnicas:

- *Random Subsampling* (submuestreo aleatorio):
 - Se repite k veces el procedimiento anterior (*hold out*) y se hace la media de los errores obtenidos
 - k experimentos tomando como conjunto de prueba diferentes subconjuntos del conjunto de datos





EVALUACIÓN

- Técnicas:

- *Bootstrapping*

- Muestreo uniforme con reemplazo de los ejemplos disponibles para entrenamiento
 - Una vez que se escoge un ejemplo, se vuelve a dejar en el conjunto y puede que se vuelva a escoger
 - El conjunto de test serán las muestras no escogidas para entrenamiento
 - **0.632 bootstrap:**
 - Dado un conjunto de d datos, se toman d muestras.
 - Los datos que no se escojan formarán parte del conjunto de prueba.
 - En torno al 63.2% de las muestras estarán en el “bootstrap” (el conjunto de entrenamiento) y el 36.8% caerá en el conjunto de prueba ya que $(1 - 1/d)^d \approx e^{-1} = 0.368$

EVALUACIÓN

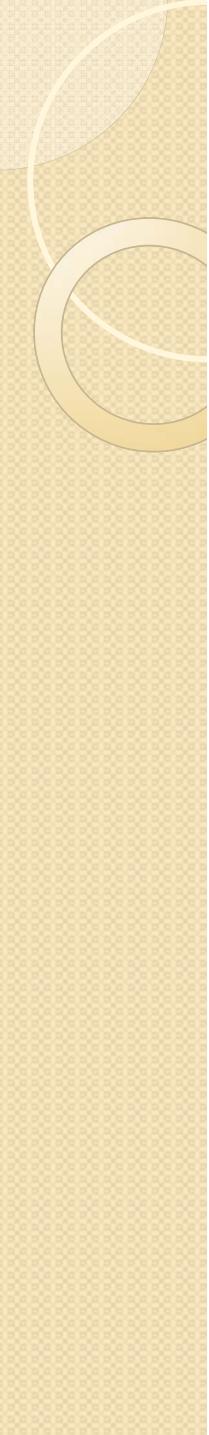
- Técnicas:
 - *Bootstrapping*

Conjunto de Datos completo	x_1	x_2	x_3	x_4	x_5			
Experimento 1	x_3	x_1	x_3	x_3	x_5	x_2	x_4	
Experimento 2	x_5	x_5	x_3	x_1	x_2		x_4	
Experimento 3	x_5	x_5	x_1	x_2	x_1		x_3	x_4
			⋮					
Experimento K	x_4	x_4	x_4	x_4	x_1	x_2	x_3	x_5
						Conjuntos de entrenamiento	Conjuntos de prueba	



EVALUACIÓN

- Técnicas:
 - *Leave k-Out*
 - Separar k datos, entrenar con los demás, y evaluar con los k datos separados
 - Repetir lo anterior apartando otros k datos diferentes
 - Cada vez se entrena y hace test con datos distintos
 - Obtener una media de los resultados
 - Relacionada con esta, la técnica de **validación cruzada**



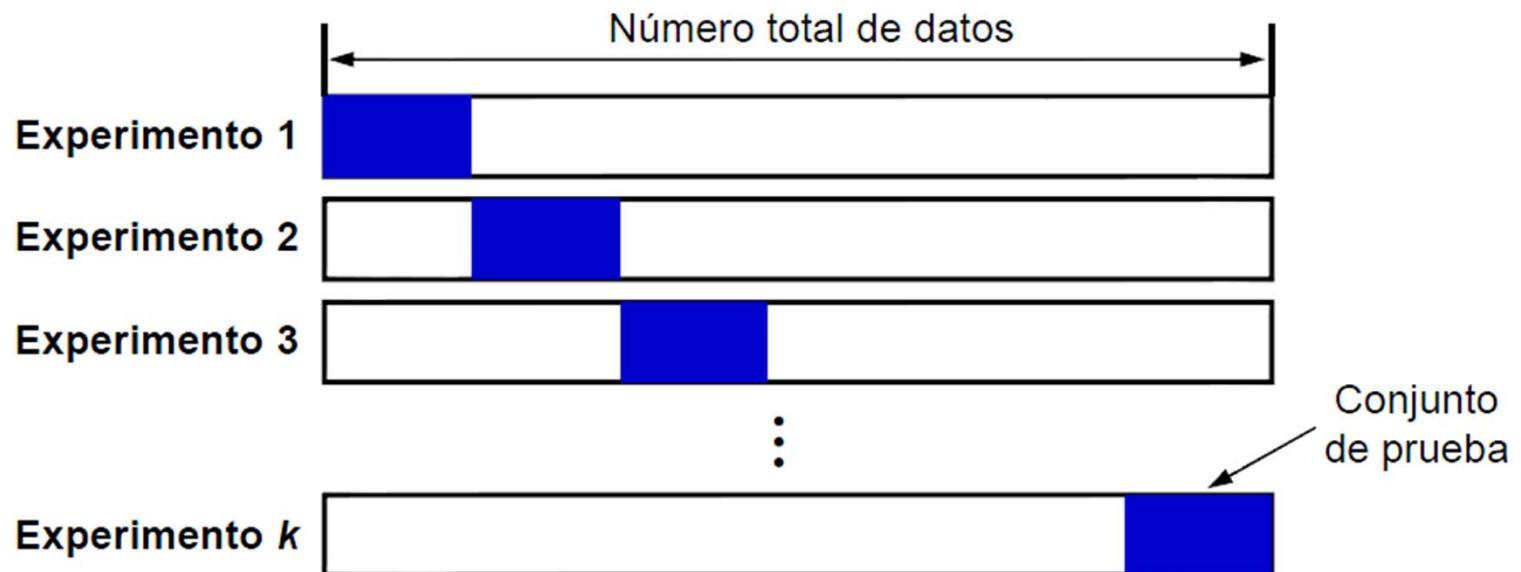
EVALUACIÓN

- Técnicas:

- Validación cruzada
 - *k-fold Cross-Validation*
 - Se divide aleatoriamente el conjunto de datos en k subconjuntos disjuntos $\{D_1, \dots, D_k\}$ del mismo tamaño
 - Típicamente, $k = 10$
 - El algoritmo es entrenado y probado k veces:
 - En la iteración i , se usa el subconjunto D_i como conjunto de prueba (test) y los $k-1$ restantes como conjunto de entrenamiento
 - El resultado será la media de los resultados de test
 - Garantiza haber probado todos los datos
 - **Es ligeramente pesimista**
 - Validación cruzada estratificada:
 - Conseguir que en cada subconjunto D_i haya igual número de ejemplares de cada clase representada

EVALUACIÓN

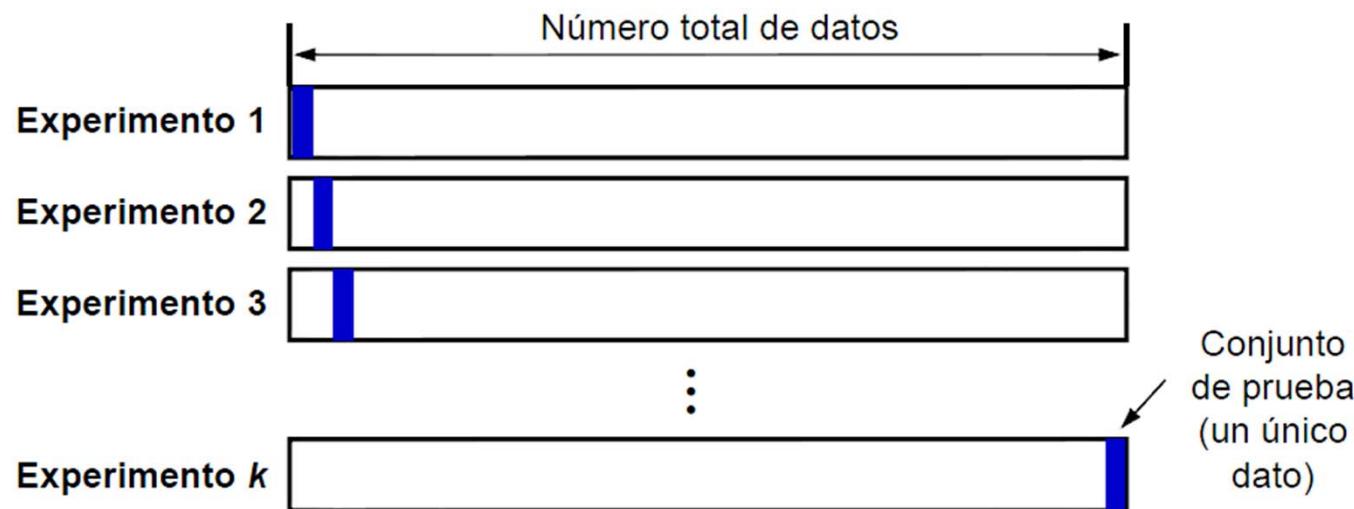
- Técnicas:
 - Validación cruzada
 - Método similar al submuestreo aleatorio
 - Ventaja de k-fold: todas las muestras del conjunto de datos se usan alguna vez para entrenar o como parte del conjunto de prueba



EVALUACIÓN

- Técnicas:

- Validación cruzada y *leave one-out*
 - Es el caso extremo de la validación cruzada *k-fold* tomando *k* como el número de muestras *N*
 - Para un conjunto de *N* muestras se realizan *N* experimentos
 - En cada experimento se emplean *N-1* datos para entrenar y el dato restante de prueba





EVALUACIÓN

- Técnicas:

- Validación cruzada:

- ¿Cuántos subconjuntos y experimentos realizar?
 - Si se elige un gran número de subconjuntos
 - El error estimado será muy preciso (sesgo pequeño respecto al error real)
 - Pero...
 - La varianza del error real será elevado
 - Tiempo computacional elevado (muchos experimentos)
 - Si se eligen pocos subconjuntos
 - Tiempo computacional reducido (pocos experimentos)
 - La varianza del estimador será pequeña
 - Pero...
 - El error estimado será menos preciso (sesgo mayor respecto al error real)

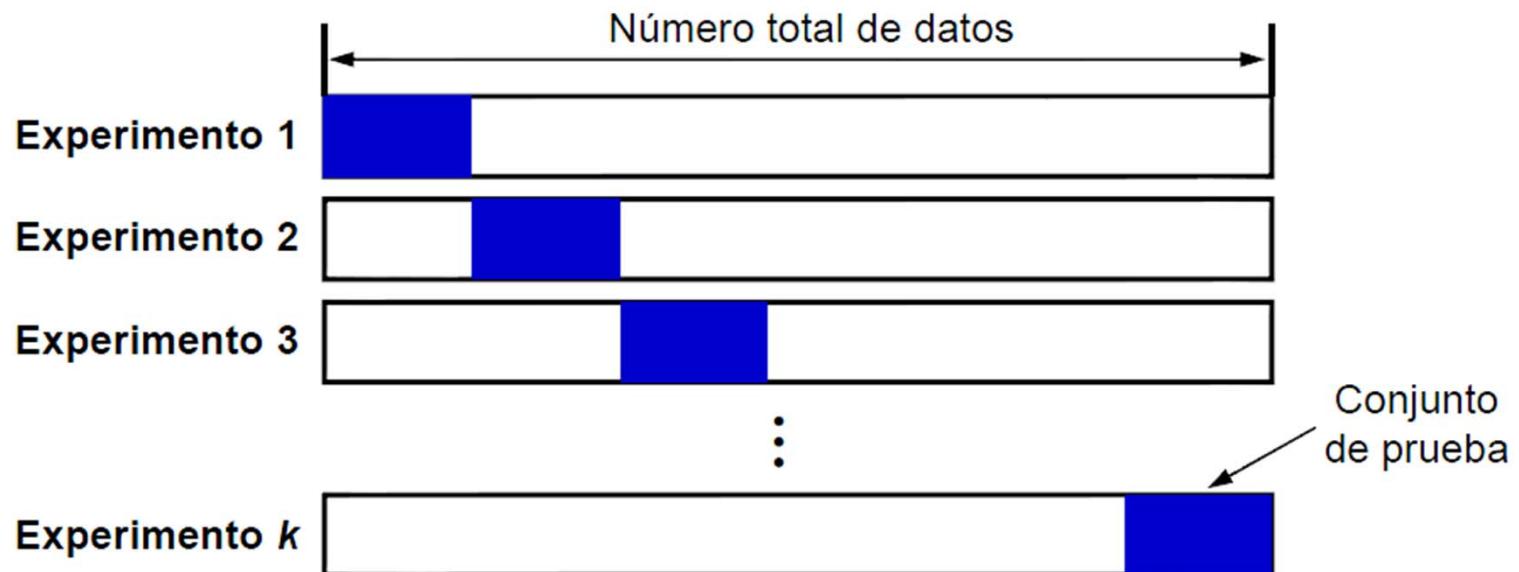


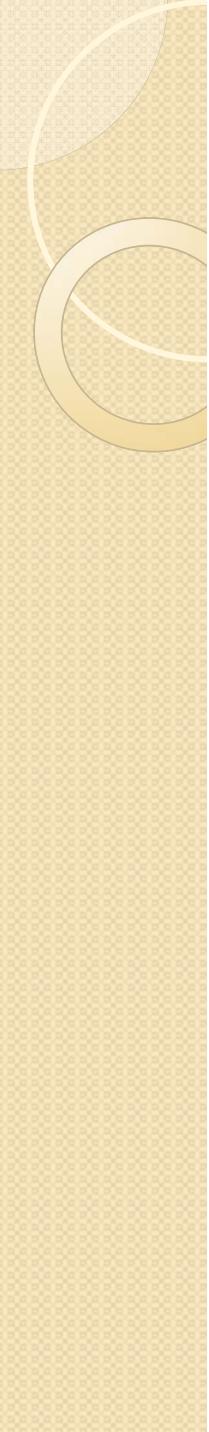
EVALUACIÓN

- Técnicas:
 - Validación cruzada:
 - En la práctica, la elección del número de subconjuntos depende del tamaño del conjunto de datos
 - Para conjuntos de datos de gran tamaño incluso una validación cruzada *3-fold* será bastante precisa
 - Para conjuntos de datos pequeños, se puede emplear la *leave one-out* para emplear como entrenamiento tantos datos como sea posible
 - Una elección habitual de la *k-fold* es $k=10$

EVALUACIÓN

- Con estas técnicas se minimiza el impacto de dividir los patrones de forma aleatoria
 - Si el algoritmo es determinístico, no hay más fuentes de variabilidad
 - En cambio, si es estocástico el resultado de cada uno de estos experimentos no será representativo:



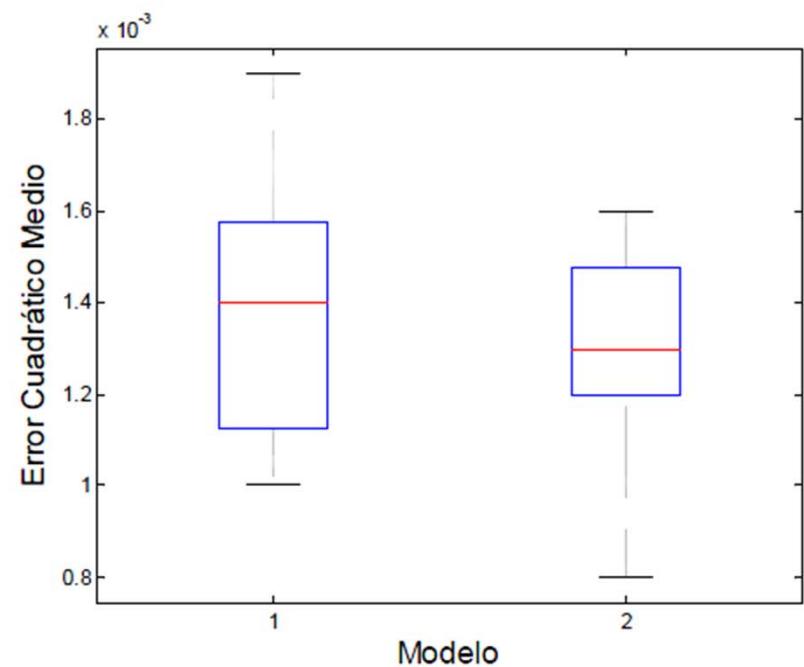
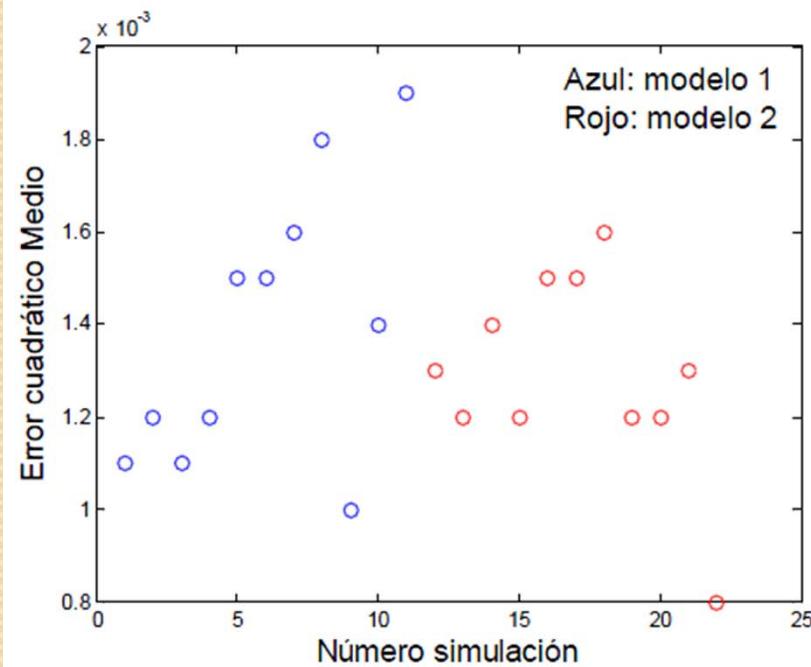


EVALUACIÓN

- Con estas técnicas se minimiza el impacto de dividir los patrones de forma aleatoria
 - Si el algoritmo es determinístico, no hay más fuentes de variabilidad
 - En cambio, si es estocástico el resultado de cada uno de estos experimentos no será representativo:
 - Cada vez que se entrene cada experimento dará un resultado distinto
 - El resultado de un único entrenamiento no es significativo
 - Necesario repetir todo el proceso (*validación cruzada, bootstrapping, leave-one-out, etc.*)
 - **Con las mismas particiones de datos de cada vez**
 - Se “elimina temporalmente” la variabilidad debida a la partición de los datos para intentar minimizar la variabilidad debida al algoritmo

EVALUACIÓN

- Para comparar dos modelos:
 - ¿Cuál es mejor en términos de error?



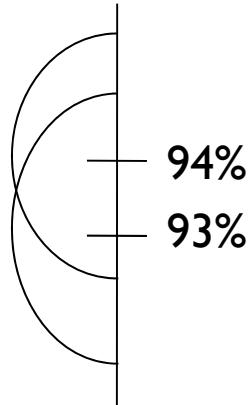


EVALUACIÓN

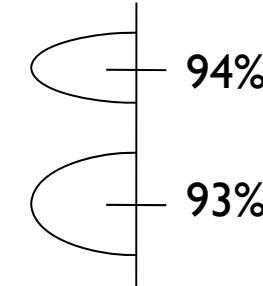
- Para comparar dos modelos:
 - Dos modelos a comparar sobre un conjunto de datos
 - Experimentos realizados siguiendo una estrategia
 - Por ejemplo: *k-fold cross-validation*
 - Conjunto de k errores cometidos por el modelo 1 (media)
 - Conjunto de k errores cometidos por el modelo 2 (media)
 - ¿El rendimiento de ambos modelos es igual? ¿Hay diferencias significativas entre ambos?
 - Comparar la media de los errores cometidos

EVALUACIÓN

- Para comparar dos modelos:
 - No basta con comparar las medias
 - Es posible que, con la desviación típica, sea

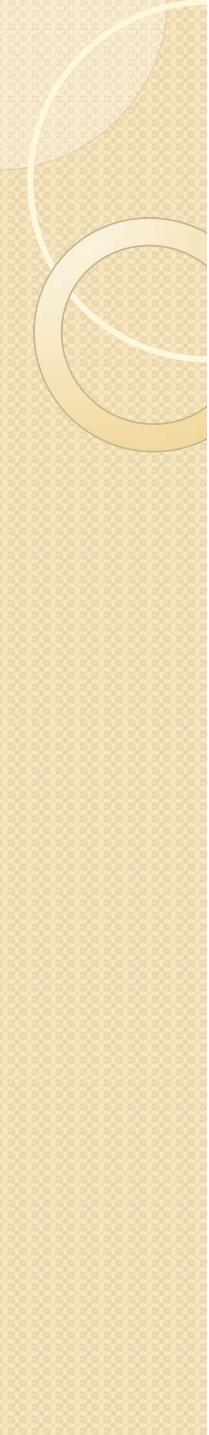


Un test estadístico podría decir
que los resultados son
estadísticamente similares



Un test estadístico diría que los
resultados son estadísticamente
distintos

- Para comparar correctamente estas medias:
 - Contraste de hipótesis
 - Método estadístico para comprobar la validez o no de una hipótesis (hipótesis nula)



EVALUACIÓN

- Comparación de dos modelos:
 - Etapas del contraste de hipótesis (opción I):
 1. Definir la hipótesis nula (H_0)
 2. Seleccionar la medida de discrepancia (estadístico del contraste o función de decisión)
 3. Elegir el nivel de significación (α) del test
 - Probabilidad de rechazar H_0 siendo cierta
 4. Determinación de la región crítica (C) o de rechazo
 5. Cálculo de la medida de discrepancia (d) para la muestra particular
 6. Resultado del test:
 - Si d está en $C \rightarrow$ Rechazar H_0
 - Si d no está en $C \rightarrow$ Aceptar H_0



EVALUACIÓN

- Comparación de dos modelos:
 - Etapas del contraste de hipótesis (opción 2):
 1. Definir la hipótesis nula (H_0)
 2. Seleccionar un test estadístico (estadístico del contraste) que pueda emplearse para evaluar la validez de H_0
 3. Elegir el nivel de significación (α) del test
 - Probabilidad de rechazar H_0 siendo cierta
 4. Calcular el p-valor
 - Probabilidad de obtener una discrepancia mayor de la observada siendo H_0 cierta
 5. Comparar el p-valor obtenido con nivel de significación:
 - Si $p \leq \alpha \rightarrow$ Rechazar H_0
 - Si $p > \alpha \rightarrow$ Aceptar H_0
 - Por tanto, el p-valor informa sobre cuál sería el nivel de significación más pequeño que nos permitiría rechazar la hipótesis nula



EVALUACIÓN

- Tipos de test estadísticos empleados en selección de modelos:
 - t-test
 - Test de Wilcoxon



EVALUACIÓN

- Tipos de test estadísticos empleados en selección de modelos:
 - t-test
 - Test paramétrico
 - Evalúa las diferencias entre las medias (errores medios) de dos modelos.
 - Hipótesis nula: $\mu_1 - \mu_2 = 0$
 - Suposiciones de este estadístico:
 - Ambas distribuciones siguen una distribución normal con idénticas varianzas

EVALUACIÓN

- Tipos de test estadísticos empleados en selección de modelos:
 - t-test
 - El intervalo de confianza para la diferencia de medias poblacionales está centrado en la diferencia de las medias muestrales, siendo sus límites superior e inferior:
 - donde $t_{\alpha/2}$ es el valor crítico correspondiente al grado de confianza $1-\alpha$ de la distribución t de Student con n_1+n_2-2 grados de libertad y S
 - es una estimación de la desviación típica común a ambas poblaciones obtenida a partir de las varianzas de las dos muestras

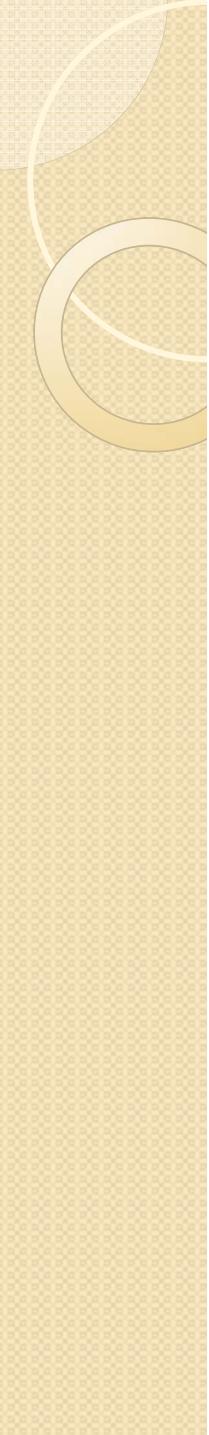
EVALUACIÓN

- Tipos de test estadísticos empleados en selección de modelos:
 - t-test
 - S_1^2 y S_2^2 son las cuasivarianzas muestrales de ambas muestras:

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2 \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2$$

- Por lo tanto, el t-test se basa en el estadístico

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$



EVALUACIÓN

- Tipos de test estadísticos empleados en selección de modelos:
 - Test de Wilcoxon
 - *Wilcoxon Signed Rank Test*
 - Test no paramétrico
 - Evalúa las diferencias entre las medianas de dos modelos
 - Hipótesis nula: $m_1 - m_2 = 0$
 - Suposiciones de este test estadístico:
 - Ninguna



EVALUACIÓN

- Tipos de test estadísticos empleados en selección de modelos:
 - Test de Wilcoxon
 - Procedimiento:
 - Se tiene n pares de observaciones (x_i, y_i) ordenadas
 - Se calculan las diferencias $z_i = y_i - x_i$
 - Se excluyen aquellos pares con $|y_i - x_i| = 0$
 - N_r es el tamaño muestral reducido
 - Se ordenan los valores absolutos $|z_1|, \dots, |z_n|$ y se les asigna su rango R_i
 - El estadístico de la prueba es
$$W = \left| \sum_{i=1}^{N_r} [sign(y_i - x_i) \cdot R_i] \right|$$
 - se consulta en tablas para determinar si se acepta o no la hipótesis nula

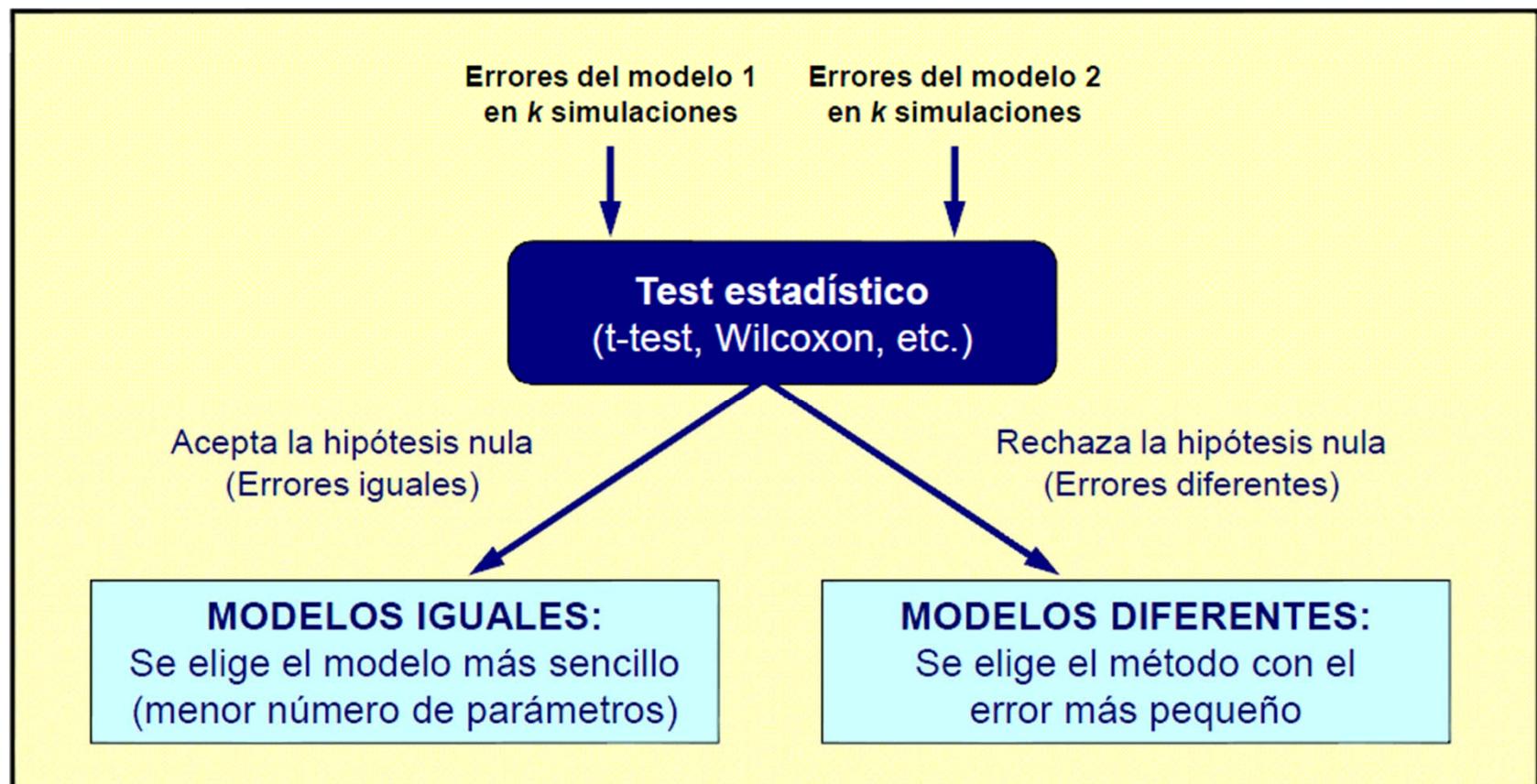


EVALUACIÓN

- ¿Cuál de los dos métodos de contraste de hipótesis es el más adecuado?
 - Si se cumplen las suposiciones del t-test, este método es más potente
 - Mayor probabilidad de rechazar H_0 cuando es falsa
 - Cuando no se cumplen las suposiciones del t-test, el de Wilcoxon es más potente y más fiable
 - No asume ninguna distribución
 - El test de Wilcoxon es más robusto frente a casos atípicos
 - En general, si no se conoce la distribución de los errores de cada método, emplear test de Wilcoxon

EVALUACIÓN

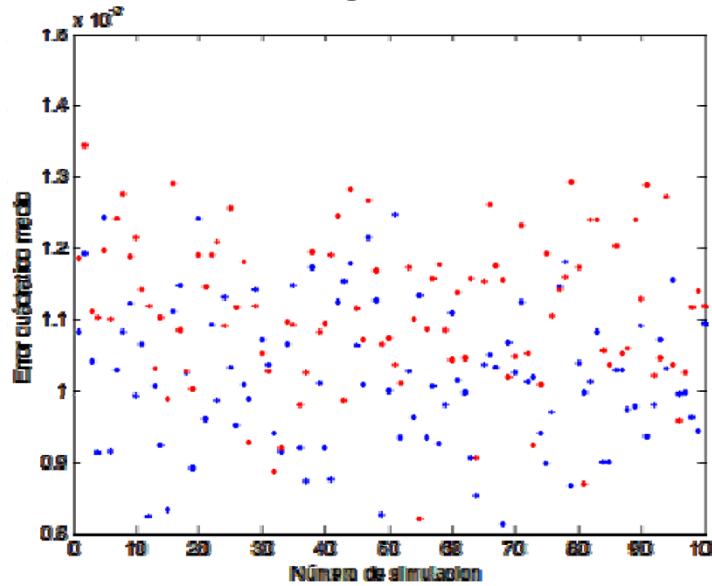
- Metodología de comparación:



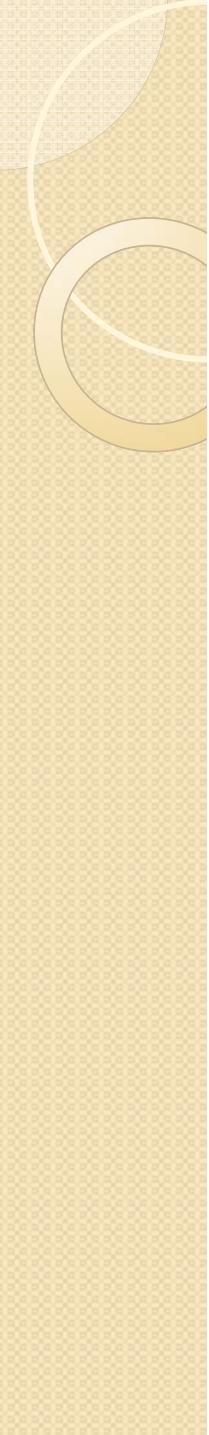
EVALUACIÓN

- Ejemplo:

- Dados dos modelos diferentes con los siguientes errores en el conjunto de prueba de 100 simulaciones (ambas muestras de errores siguen una distribución normal)



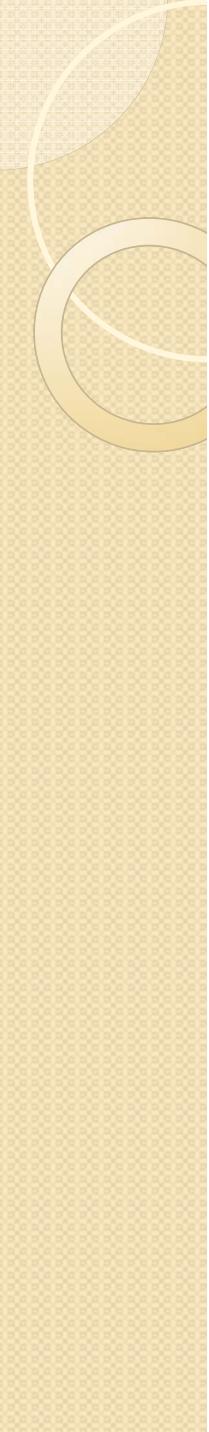
- ¿Hay diferencias estadísticamente significativas entre ambos modelos?



EVALUACIÓN

- Ejemplo:

- Se realiza un t-test para comprobarlo:
 - Hipótesis nula (H_0): $\mu_1 - \mu_2 = 0$
 - Ambas medias son iguales
 - Nivel de significación: 0.01
 - Probabilidad de rechazar la hipótesis nula siendo cierta
 - Resultado obtenido:
 - p-valor del test: $1.1934e^{-9}$
 - Puesto que $p \leq \alpha$
 - Rechazamos H_0 con un nivel de confianza del 99%



EVALUACIÓN

- Para comparar 3 o más modelos:
 - No es apropiado simplemente comparar cada par de modelos empleando por ejemplo un t-test:
 - La posibilidad de detectar incorrectamente una diferencia significativa aumenta con el número de comparaciones
 - En este caso:
 - Emplear un análisis de varianza ANOVA o un test de Kruskal-Wallis para identificar si hay una diferencia significativa entre todas las medias
 - Si el test de varianza concluye que sí hay diferencias
 - Hay que investigar cuáles son diferentes empleando un método de comparación múltiple
 - Si el test de varianza concluye que no hay diferencias
 - Todas las medias iguales -> todos los modelos iguales
 - Se elige el más sencillo



EVALUACIÓN

- Comparación de múltiples modelos:
 - ANOVA:
 - Test paramétrico que compara las medias de diversos modelos
 - Hipótesis nula: todas las medias son iguales
 - Provienen de la misma población o de diferentes poblaciones pero con la misma media
 - Suposiciones del test:
 1. Todas las muestras de las diferentes poblaciones están normalmente distribuidas
 2. Todas las muestras de las diferentes poblaciones tienen la misma varianza
 3. Todas las observaciones son mutuamente independientes
 - El test sigue siendo robusto para observaciones que no cumplen “ligeramente” las dos primeras suposiciones



EVALUACIÓN

- Comparación de múltiples modelos:
 - Kruskal-Wallis:
 - Test no paramétrico que compara las medias de diversos modelos
 - Hipótesis nula: todas las medias son iguales
 - Provienen de la misma población o de diferentes poblaciones pero con la misma media
 - Suposiciones del test:
 - Todas las poblaciones provienen de una población continua
 - Todas las observaciones son mutuamente independientes

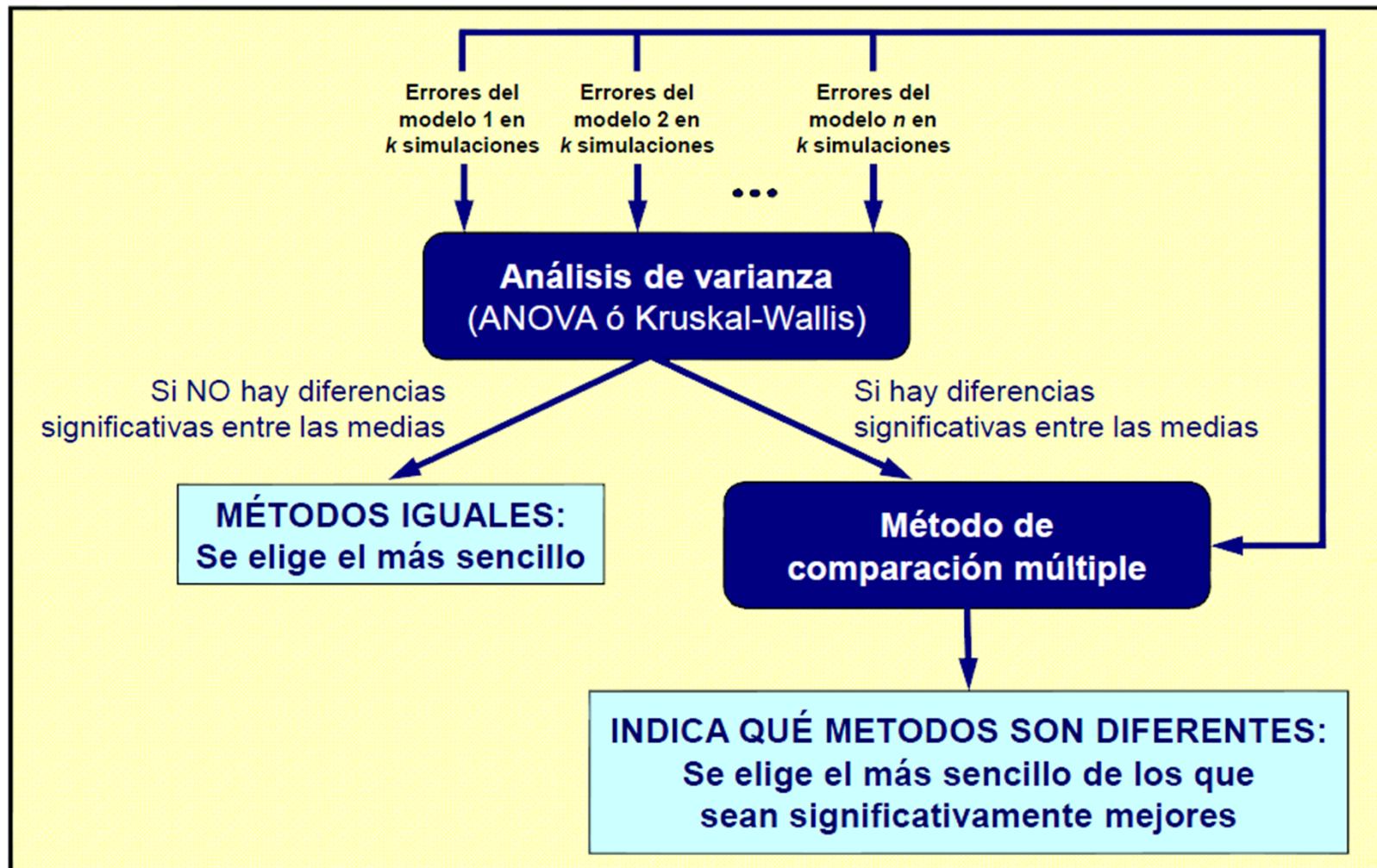


EVALUACIÓN

- Comparación de múltiples modelos:
 - Comparación de las diferencias entre cada par de medias con ajustes apropiados a la comparación múltiple:
 - Método de Tukey
 - Método de Scheffé
 - Método de Bonferroni

EVALUACIÓN

- Comparación de múltiples modelos:





EVALUACIÓN

- Al aplicar determinados modelos para resolver un problema es necesario configurarlos
 - Encontrar la mejor configuración del modelo
 - Ejemplos:
 - qué arquitectura es mejor en una RNA
 - qué kernel es mejor en un SVM
 - Probar (entrenar) distintas configuraciones y, al mismo tiempo, estimar el error real del sistema final
 - Evaluar la mejor configuración con un conjunto de test



EVALUACIÓN

- Si es necesario configurar el modelo, son necesarios tres conjuntos de datos:
 - Entrenamiento:
 - Empleado para el entrenamiento del sistema (obtención de los parámetros óptimos)
 - Validación:
 - Empleado para determinar la mejor configuración del modelo
 - Ejemplo: Número de capas en RR.NN.AA, regularización, parada temprana, etc.
 - Prueba:
 - Empleado SOLO para estimar el error real del mejor modelo obtenido con los conjuntos anteriores
- ¿Por qué utilizar conjuntos de validación y prueba independientes?:
 - Para evitar que el error esté sesgado
 - Si se mide el error en el conjunto de validación
 - Menor que el error real → optimista

EVALUACIÓN

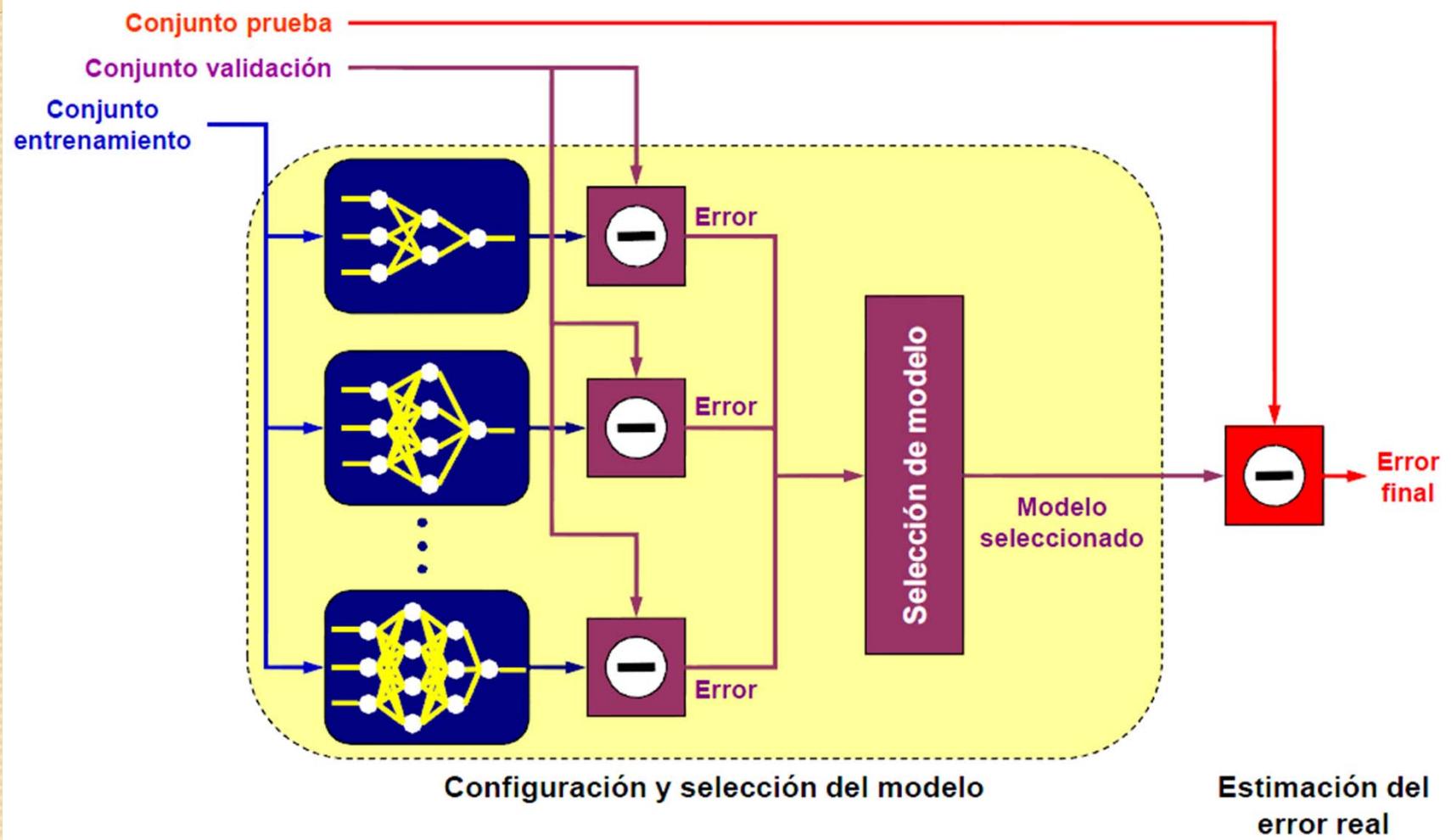
- Procedimiento experimental:

1. Dividir los datos en conjunto de entrenamiento, validación y prueba (empleando validación cruzada, *bootstrap*, etc.)
2. Seleccionar la arquitectura y los parámetros del sistema

Repetir según método	<ol style="list-style-type: none">3. Entrenar el modelo usando el conjunto de entrenamiento4. Evaluar el modelo usando el conjunto de validación
	<ol style="list-style-type: none">5. Repetir los pasos 2 a 4 usando diferentes configuraciones y parámetros iniciales6. Seleccionar el mejor modelo (la mejor configuración)7. Evaluar el modelo final empleando el conjunto de prueba

EVALUACIÓN

- Diseño experimental con 3 conjuntos de datos:

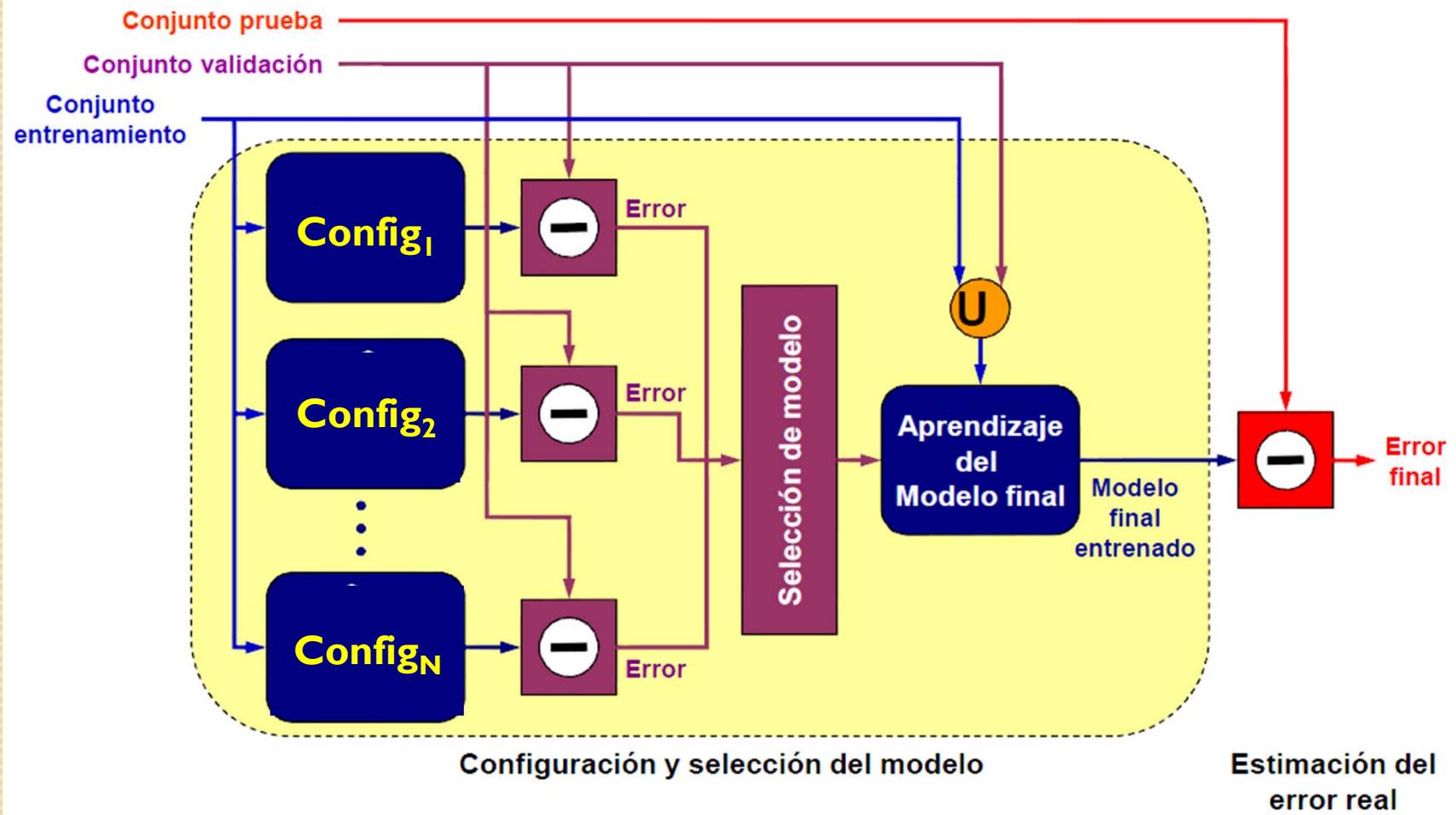


EVALUACIÓN

- Procedimiento experimental (opción 2):
 1. Dividir los datos en conjunto de entrenamiento, validación y prueba (empleando validación cruzada, bootstrap, etc.)
 2. Seleccionar la arquitectura y los parámetros del sistema
 3. Entrenar el modelo usando el conjunto de entrenamiento
 4. Evaluar el modelo usando el conjunto de validación
 5. Repetir los pasos 2 a 4 usando diferentes configuraciones y parámetros iniciales
 6. Seleccionar el mejor modelo y entrenarlo empleando el conjunto **de entrenamiento y el de validación**
 7. Evaluar el modelo final empleando el conjunto de prueba

EVALUACIÓN

- Diseño experimental con 3 conjuntos de datos:



EVALUACIÓN

- ¿Cuándo usar una u otra?
- Depende del proceso de entrenamiento:
 - Por ejemplo, con RR.NN.AA.:
 - Diferentes configuraciones: diferentes arquitecturas:



- El entrenamiento es un proceso no determinista
 - Necesario entrenar (y validar) cada arquitectura varias veces
 - Se valida a la vez que se entrena
 - Diferentes versiones de cada RNA:
 - $\text{RNA}_{\text{A}1}, \text{RNA}_{\text{A}2}, \text{RNA}_{\text{A}3}, \text{RNA}_{\text{B}1}, \text{RNA}_{\text{B}2}, \text{RNA}_{\text{B}3}, \text{RNA}_{\text{C}1}$, etc.
- De todas esas RR.NN.AA. entrenadas, se escoge la que tenga menor error de validación
 - No se entrena otra vez con el conjunto de entrenamiento+validación porque habría que dividir los patrones otra vez en entrenamiento y validación
 - El proceso de validación es algo intrínseco en el entrenamiento de una RNA
- A esa RNA escogida se le aplica el conjunto de test



EVALUACIÓN

- Por ejemplo, con SVM:
 - Diferentes configuraciones:
 - Diferentes kernel y valores de sus parámetros y C
 - El entrenamiento es un proceso determinista
 - Necesario entrenar cada configuración una única vez
 - De todas esas configuraciones entrenadas, se escoge la que tenga menor error de validación
 - Para mejorar el comportamiento del SVM, se entrena de nuevo con el conjunto de entrenamiento+validación
 - Se entrena con más patrones que antes → se espera que actúe mejor
 - En una RNA no se obtendría mejora puesto que habría que dividir en entrenamiento y validación.
 - A esa SVM escogida se le aplica el conjunto de test