

# Trade-off Sesgo - Varianza

Conjunto de  $n$  datos  $D = \{(x_1, y_1) \dots, (x_n, y_n)\}$

Relación de Dependencia  $y = f(x) + \varepsilon$

Modelo Predictivo

$$\hat{f}_k(x; D)$$

Parámetro de  
Complejidad del Modelo

$$E(\varepsilon) = 0$$
$$Var(\varepsilon) = \sigma^2$$

$$E_D \left[ (y - \hat{f}(x; D))^2 \right] = \left( \text{Bias}_D [\hat{f}(x; D)] \right)^2 + \text{Var}_D [\hat{f}(x; D)] + \sigma^2$$

Considerando los  
datos aleatorios  
provenientes de la  
conjunta  $P(x, y)$

Considerando un  
nuevo dato fijo

$$E_D [\hat{f}(x; D)] - f(x)$$

# Demostración

$$\hat{f}(x; D)$$

$$f(x)$$

$$\begin{aligned} E[(y - \hat{f})^2] &= E[(f + \varepsilon - \hat{f})^2] \\ &= E[(f + \varepsilon - \hat{f} + E[\hat{f}] - E[\hat{f}])^2] \\ \text{Desarrollando el cuadrado} \rightarrow &= E[(f - E[\hat{f}])^2 + E[\varepsilon^2] + E[(E[\hat{f}] - \hat{f})^2] + 2E[(f - E[\hat{f}])\varepsilon] + 2E[\varepsilon(E[\hat{f}] - \hat{f})] + 2E[(E[\hat{f}] - \hat{f})(f - E[\hat{f}])] \\ &= (f - E[\hat{f}])^2 + E[\varepsilon^2] + E[(E[\hat{f}] - \hat{f})^2] + 2(f - E[\hat{f}])E[\varepsilon] + 2E[\varepsilon]E[E[\hat{f}] - \hat{f}] + 2E[E[\hat{f}] - \hat{f}](f - E[\hat{f}]) \\ &= (f - E[\hat{f}])^2 + E[\varepsilon^2] + E[(E[\hat{f}] - \hat{f})^2] \\ &= (f - E[\hat{f}])^2 + \text{Var}[\varepsilon] + \text{Var}[\hat{f}] \\ \text{Es fija} \rightarrow &= \text{Bias}[\hat{f}]^2 + \text{Var}[\varepsilon] + \text{Var}[\hat{f}] \\ &= \text{Bias}[\hat{f}]^2 + \sigma^2 + \text{Var}[\hat{f}] \end{aligned}$$

Vale 0

Aditividad  
esperanza

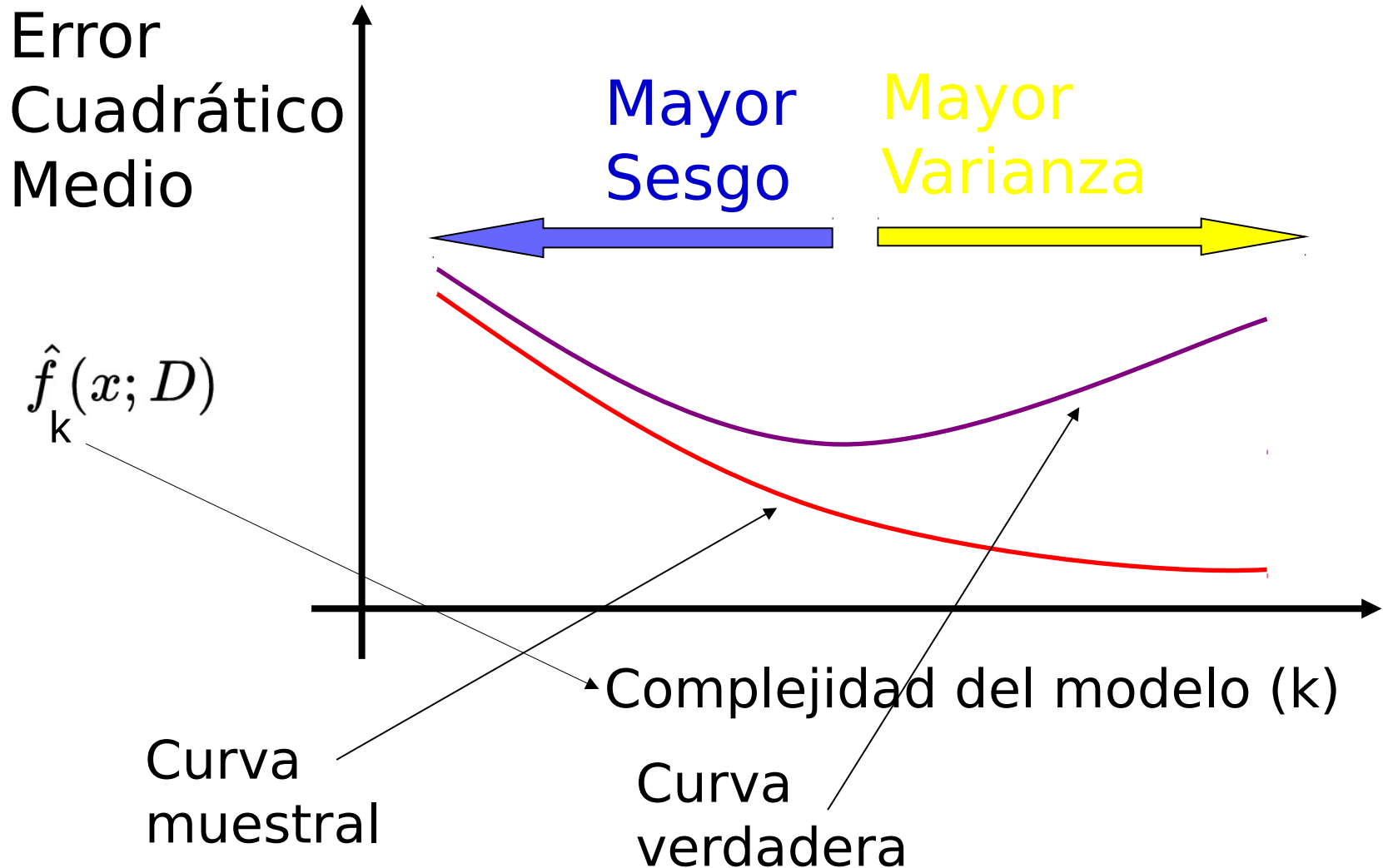
Pensando aleatoria a  
la nueva  $x \sim P$

$$\text{Var}[X] = E[X^2] - (E[X])^2$$

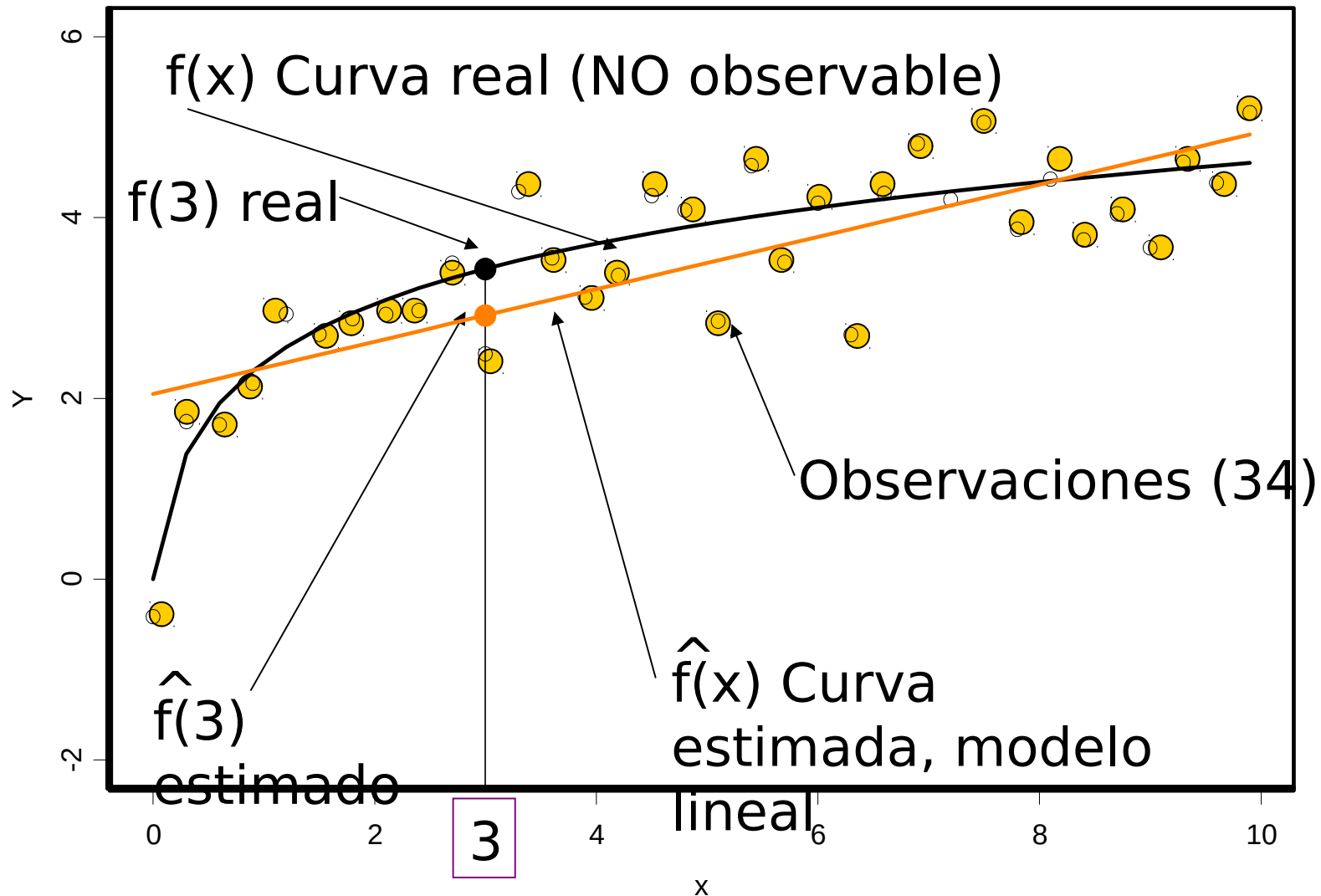
$$\text{MSE} = E_x \left\{ \text{Bias}_D[\hat{f}(x; D)]^2 + \text{Var}_D[\hat{f}(x; D)] \right\} + \sigma^2$$

# Tradeoff Sesgo - Varianza

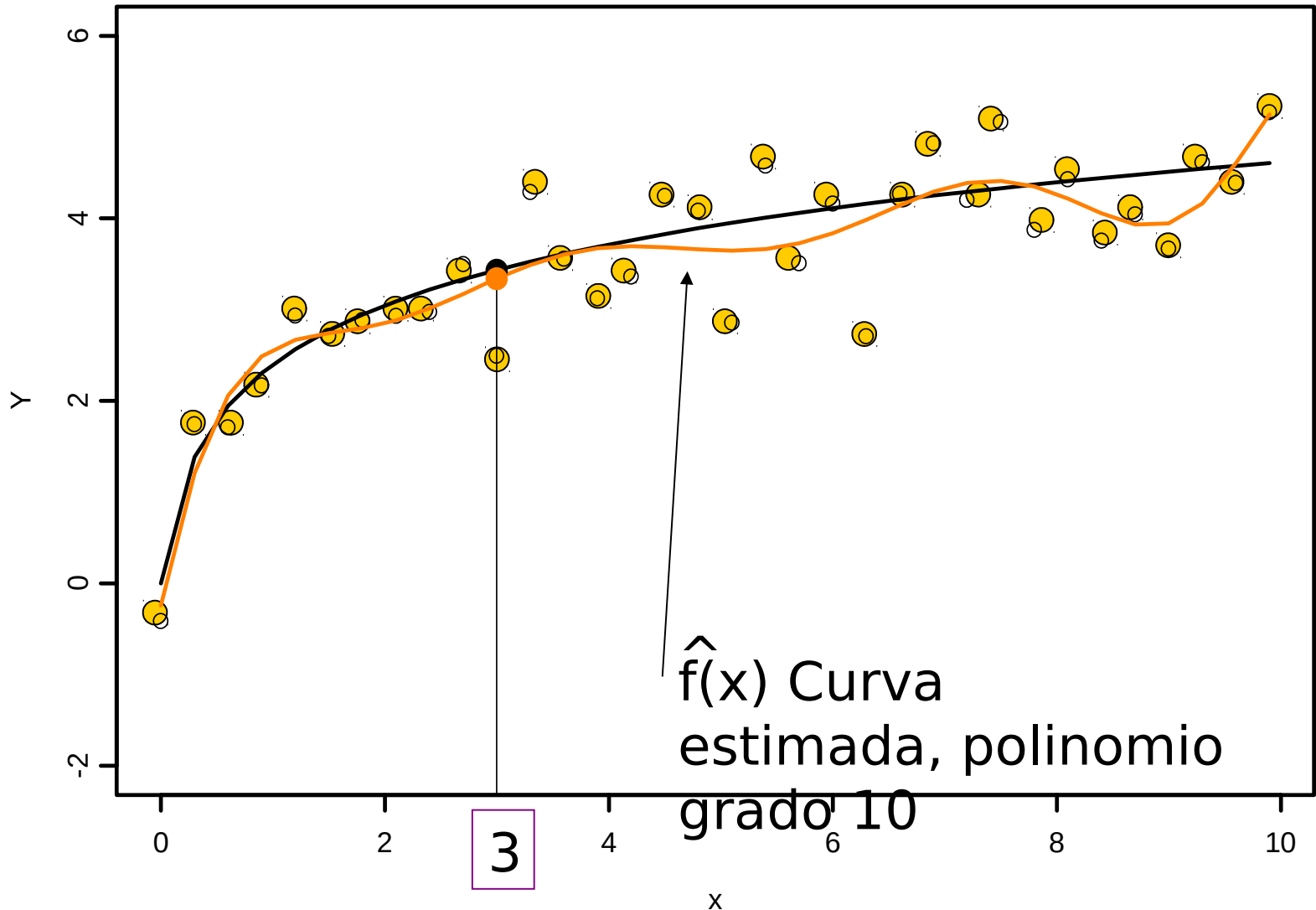
$$\text{MSE} = E_x \left\{ \text{Bias}_D [\hat{f}_k(x; D)]^2 + \text{Var}_D [\hat{f}_k(x; D)] \right\} + \sigma^2$$



$$\text{Ajuste de } Y = f(X) + \varepsilon = \ln(10 \cdot X + 1) + \varepsilon$$



$$\text{Ajuste de } Y = f(X) + \varepsilon = \ln(10 \cdot X + 1) + \varepsilon$$

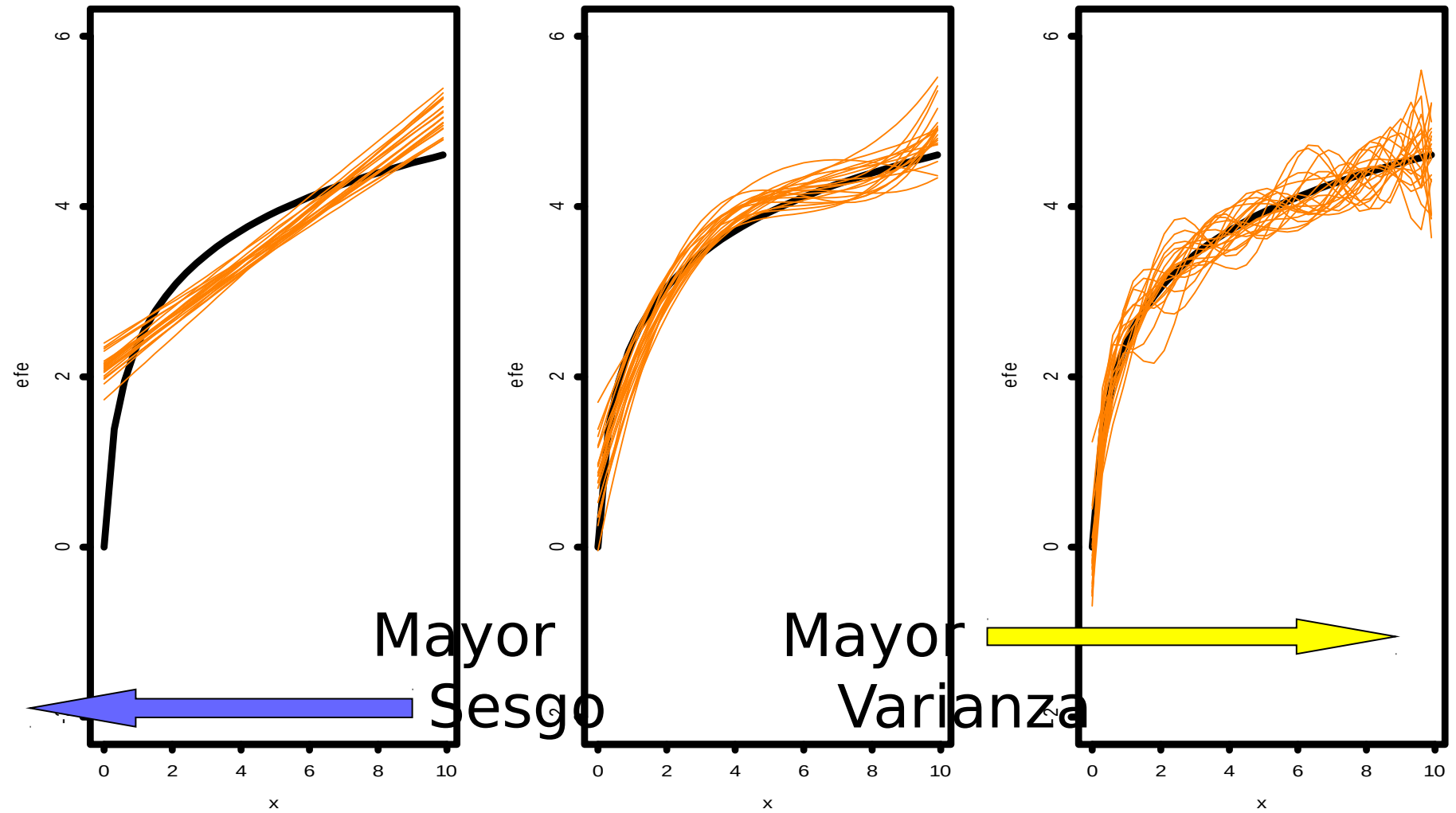


# Tradeoff Sesgo-Varianza

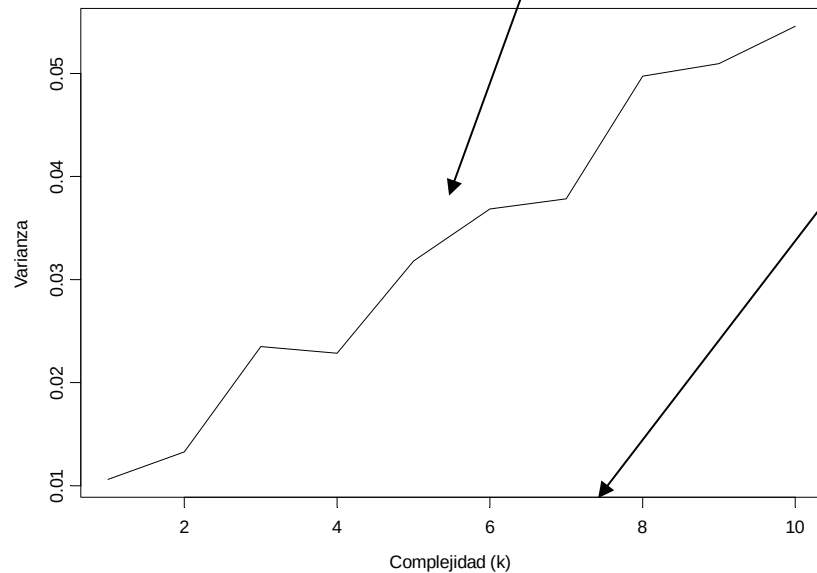
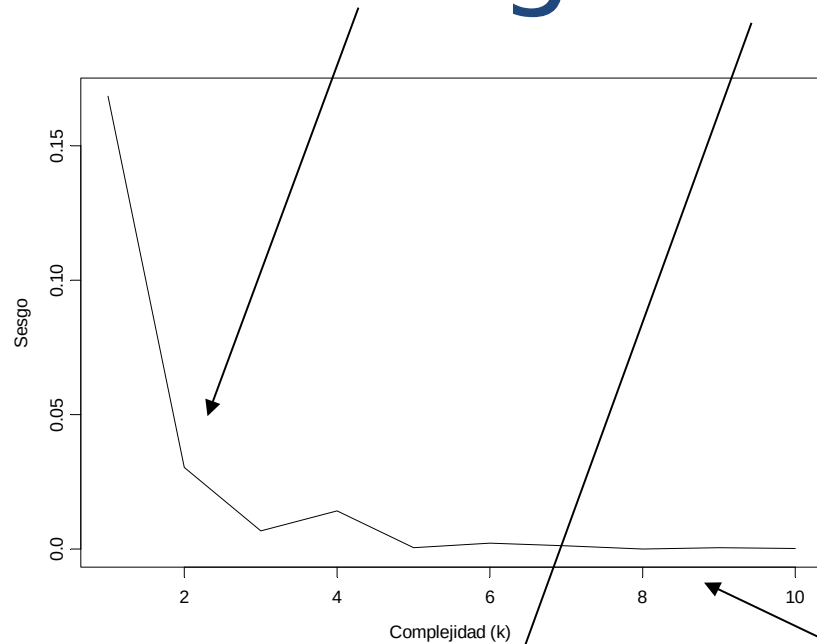
Complejidad  $k=2$   
Polinomio grado 1

Complejidad  $k=4$   
Polinomio grado 3

Complejidad  $k=11$   
Polinomio grado 10

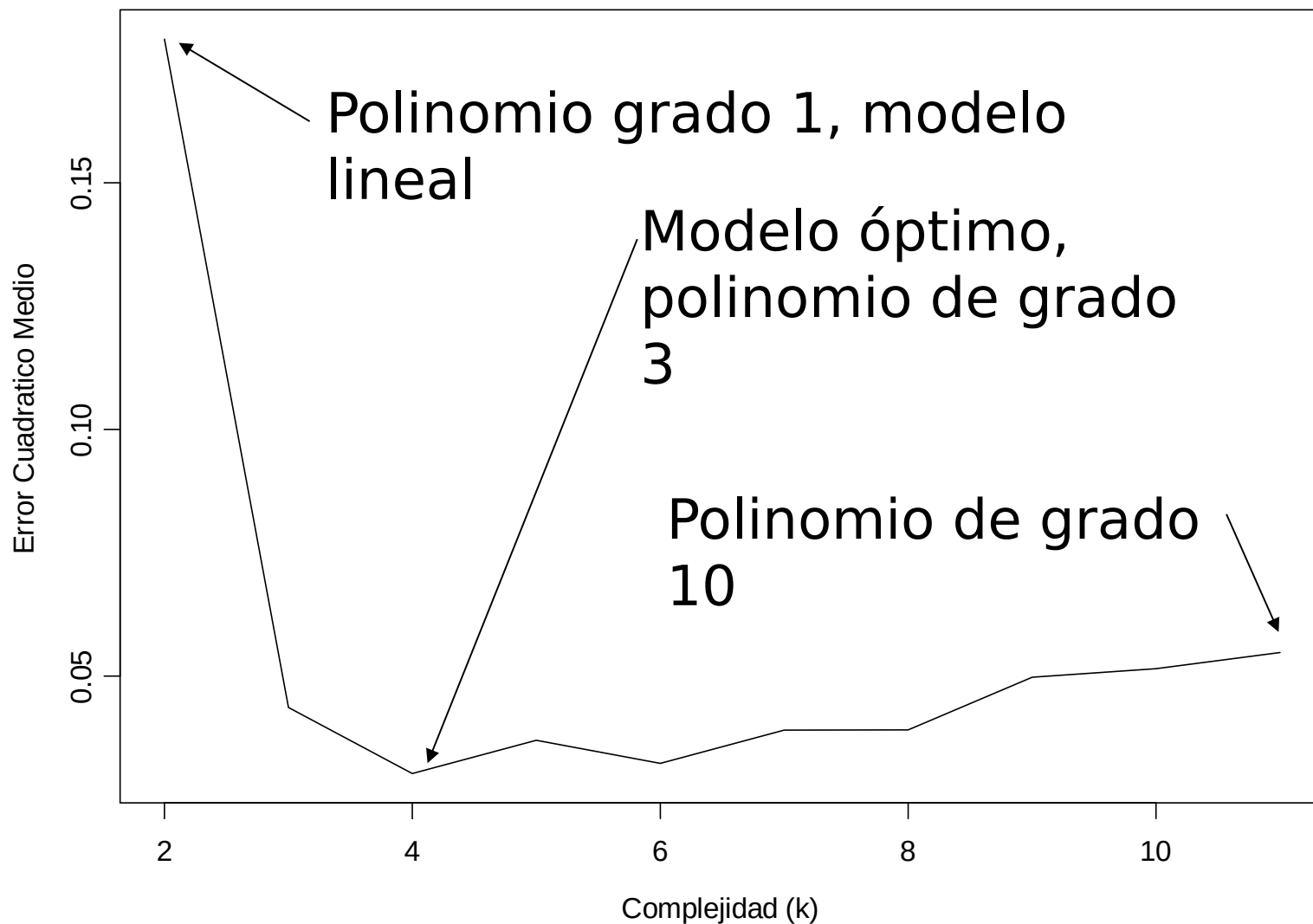


# Tradeoff Sesgo - Varianza



Complejidad del  
Modelo = Grados  
del Polinomio

# Curva de Error Cuadrático Medio





# Overfitting (sobreajuste)

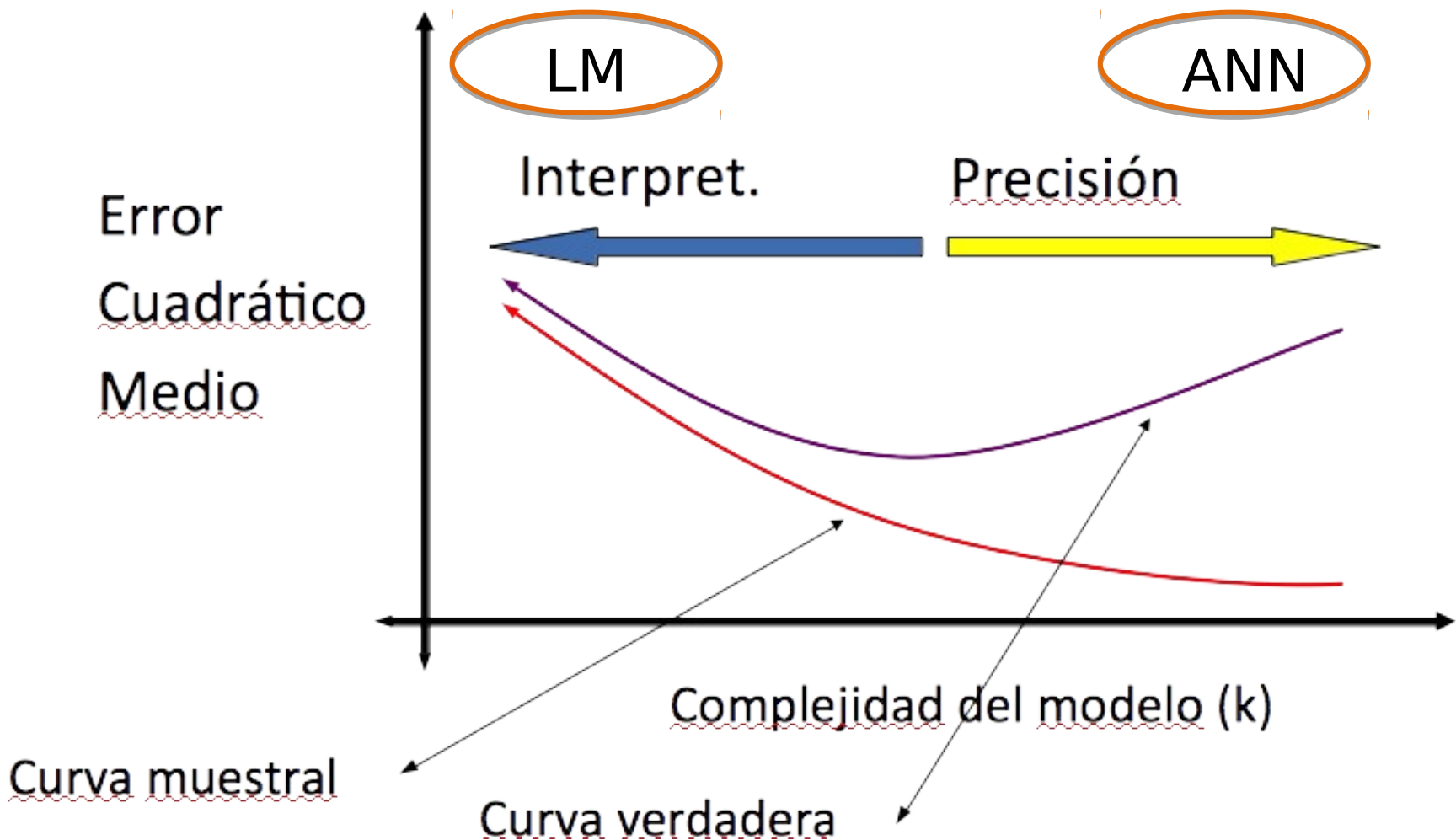
- Efecto nocivo que proviene de ajustar modelos con mas complejidad (menos parsimonia) que la que la cantidad de información muestral admite.
- Resultado de la **ALTA VARIABILIDAD** del estimador.
- Es MUCHO mas común que el SUBAJUSTE.
- Estamos genéticamente Programados para el OVERFITTING !



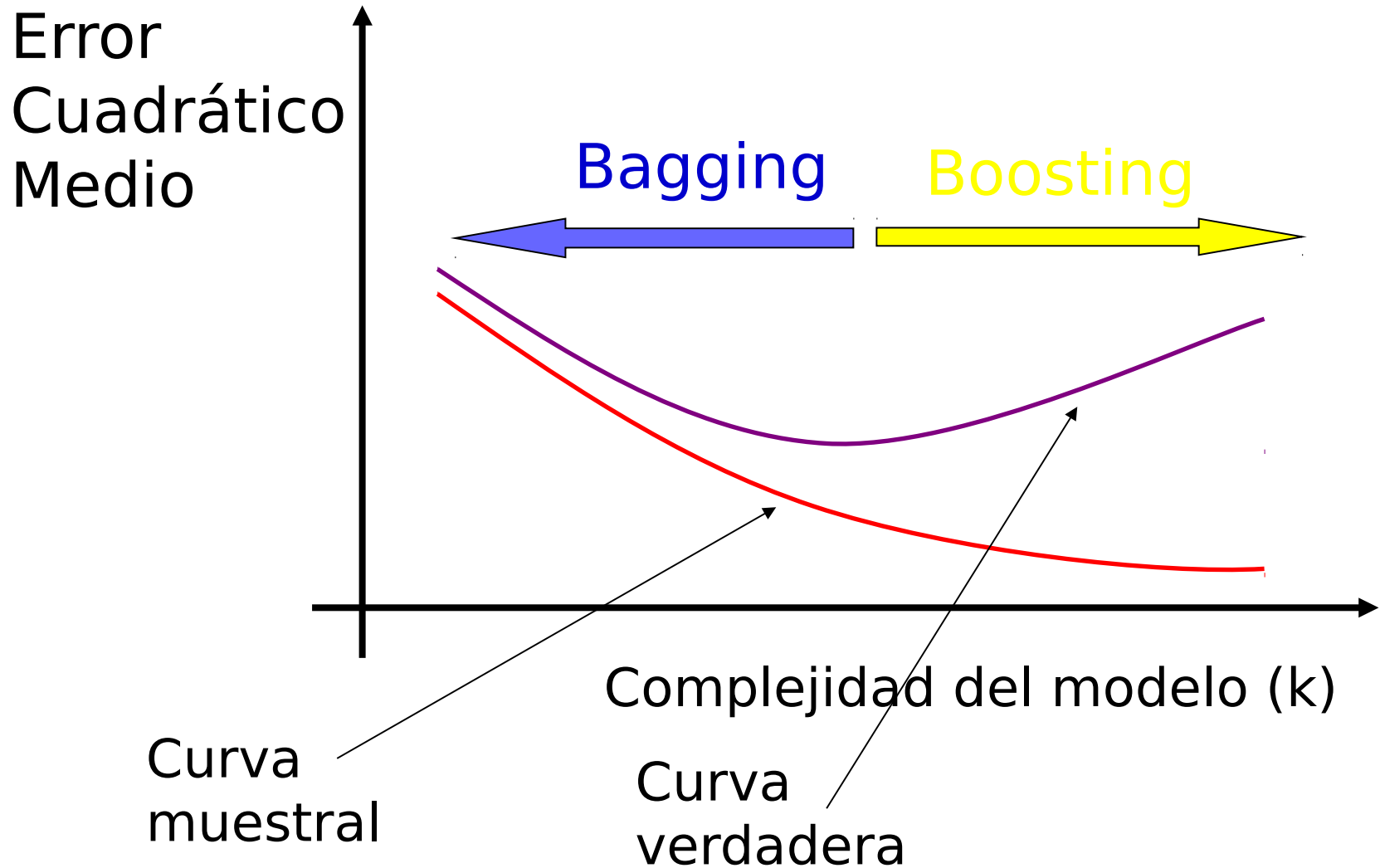
# Solución al Overfitting

- Usar modelos poco complejos (mas parsimoniosos)
- Partir la muestra: Entrenamiento / Testeo
- Usar validación cruzada
- Bootstrap y Bagging
- Usar técnicas de “Shrinkage”, como:
  - Ridge Regression
  - LASSO Regression
  - Penalización o Regularización

# Tradeoff Precisión - Interpretabilidad



# Meta - Heurísticas

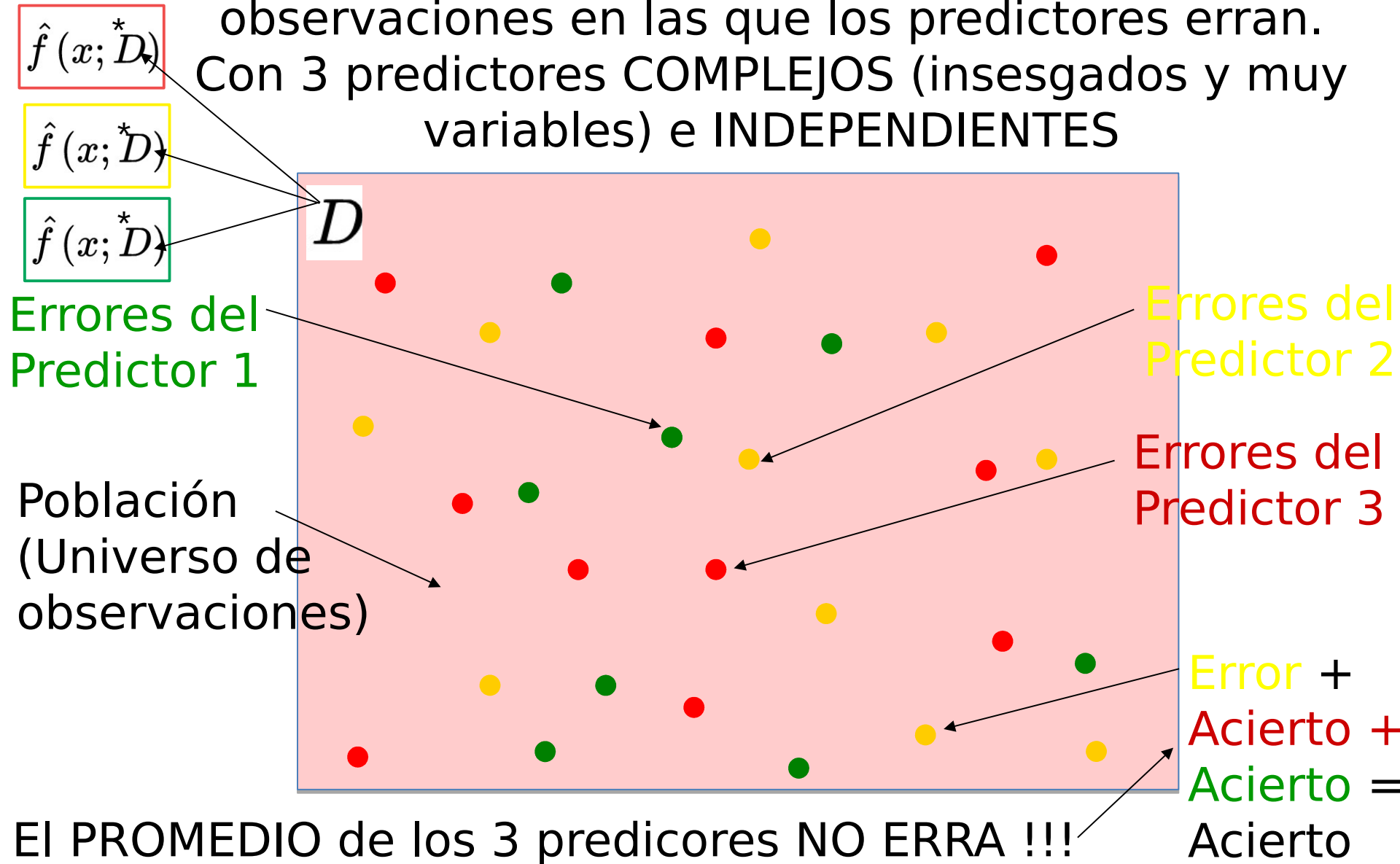


# Bagging (Bootstrap Aggregating)

- Se basa en **PROMEDIAR** los resultados de iterar la aplicación de modelos **COMPLEJOS**, “**bootstrapeando**” la muestra de entrenamiento.
- Esta técnica reduce la **VARIANZA** típico de los modelos **COMPLEJOS**.

# Intuición del Bagging en Clasificación

Los círculos de colores representan a las observaciones en las que los predictores erran.  
Con 3 predictores COMPLEJOS (insesgados y muy variables) e INDEPENDIENTES



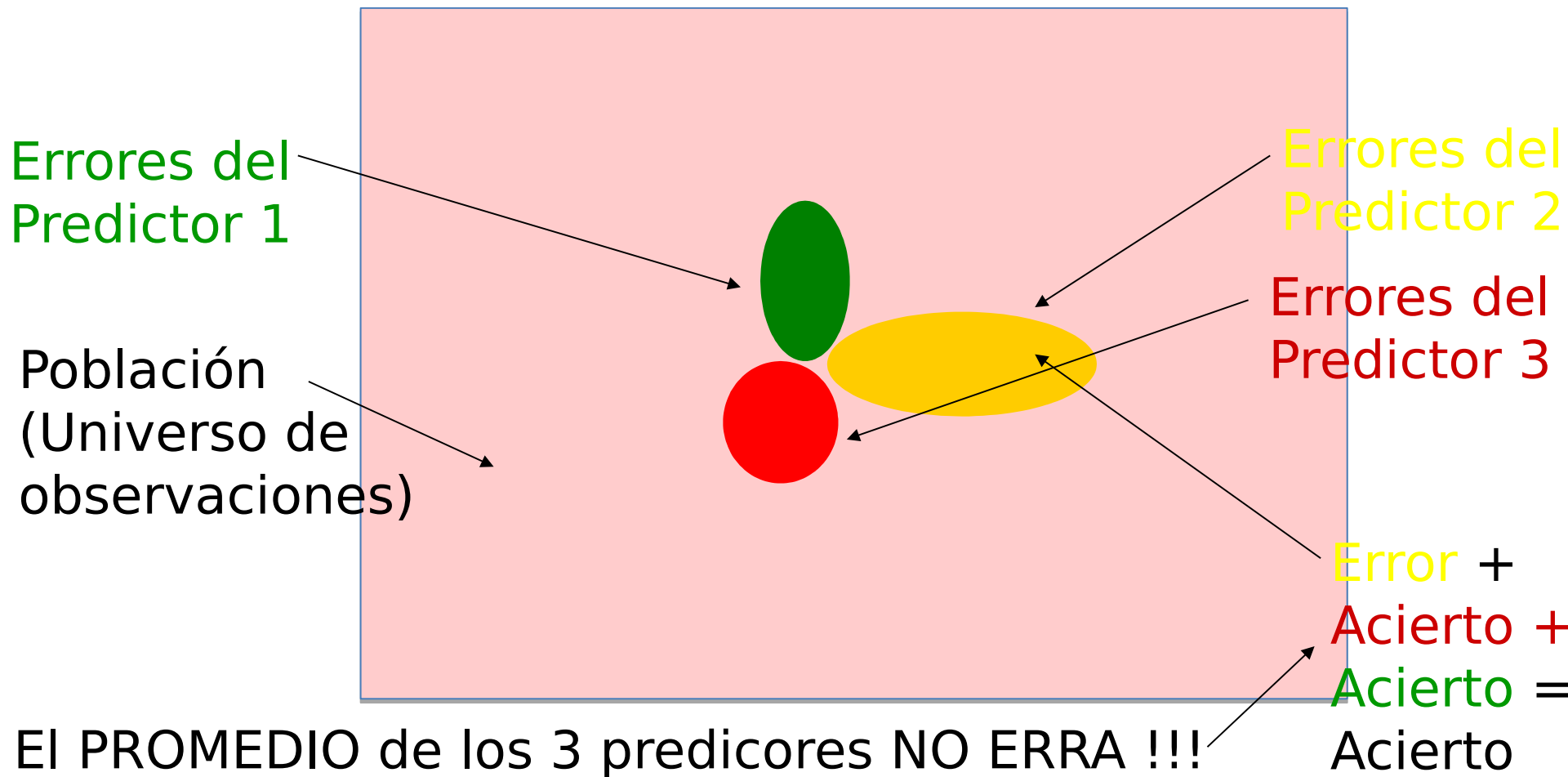
# Boosting

- Se basa en **RE-ENTRENAR** iterativamente modelos **SIMPLES** aumentando la ponderación de las observaciones PEOR predichas.
- Esta técnica reduce el **SESGO** típico de los modelos **SIMPLES**.

# Intuición del Boosting en Clasificación

Los círculos de colores representan a las observaciones en las que los predictores erran.

Con 3 predictores SIMPLES (sesgados y poco variables) e INDEPENDIENTES



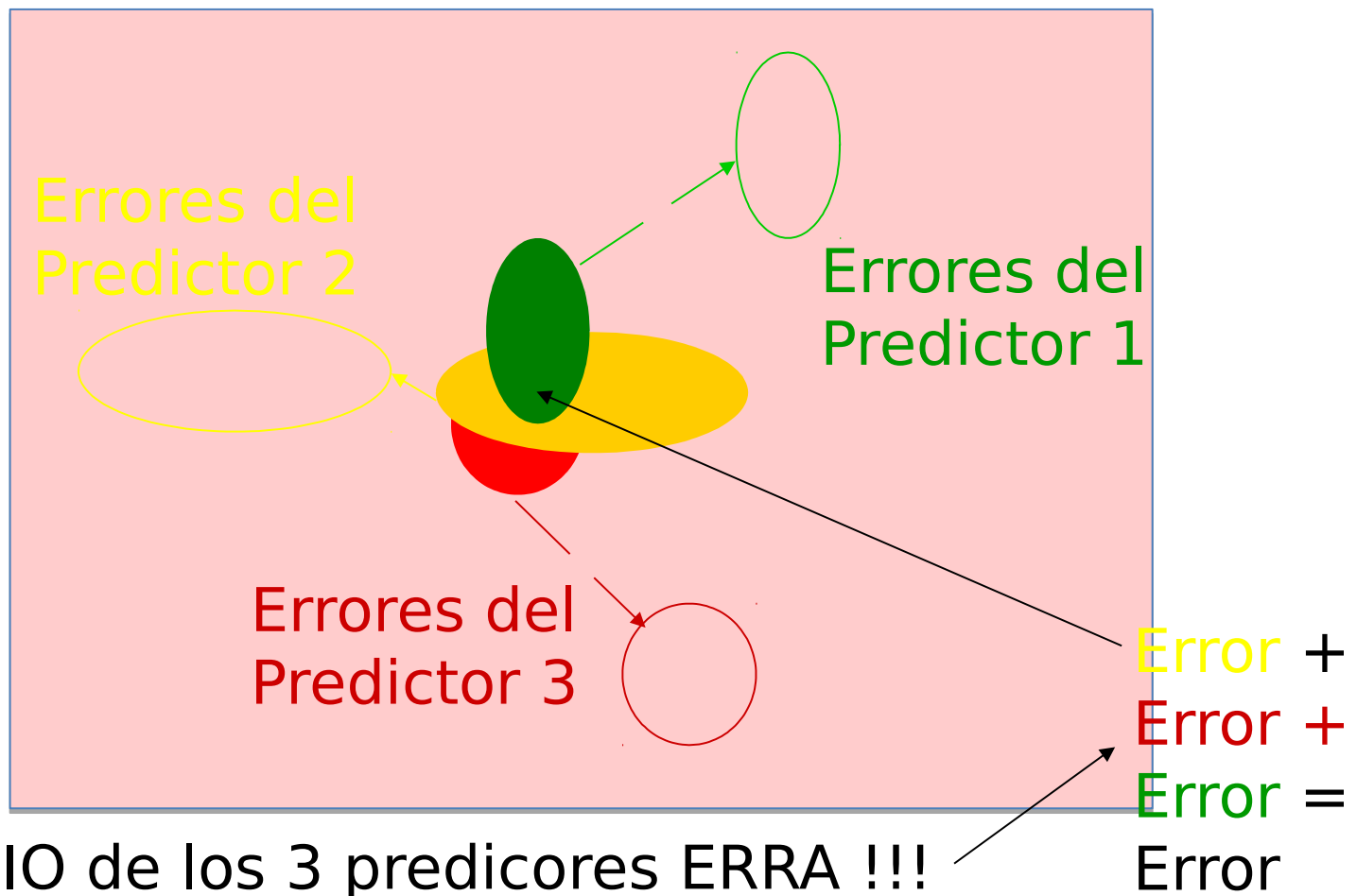


# En la Práctica: Boosting en Clasificación

Con 3 predictores SIMPLES (sesgados y poco variables) que

NO SON INDEPENDIENTES

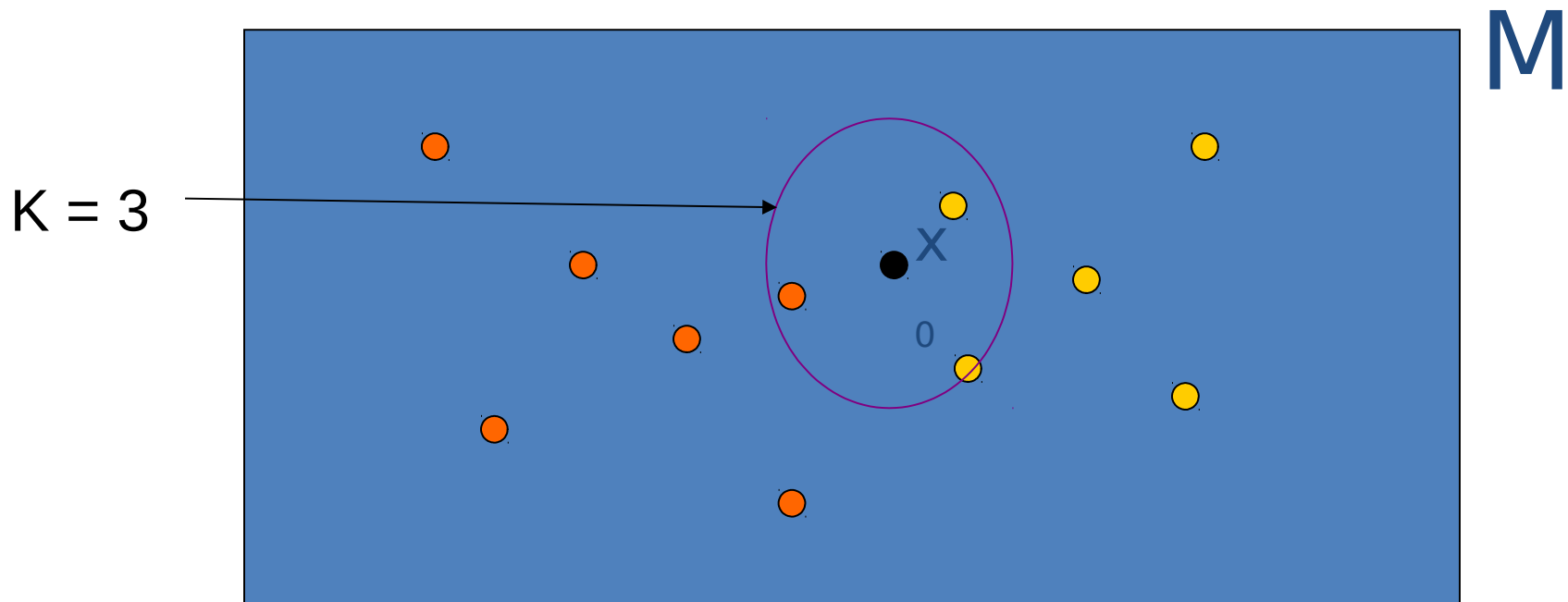
¿ Como SEPARAMOS las regiones de ERROR de los predictores?



El PROMEDIO de los 3 predicores ERRA !!!

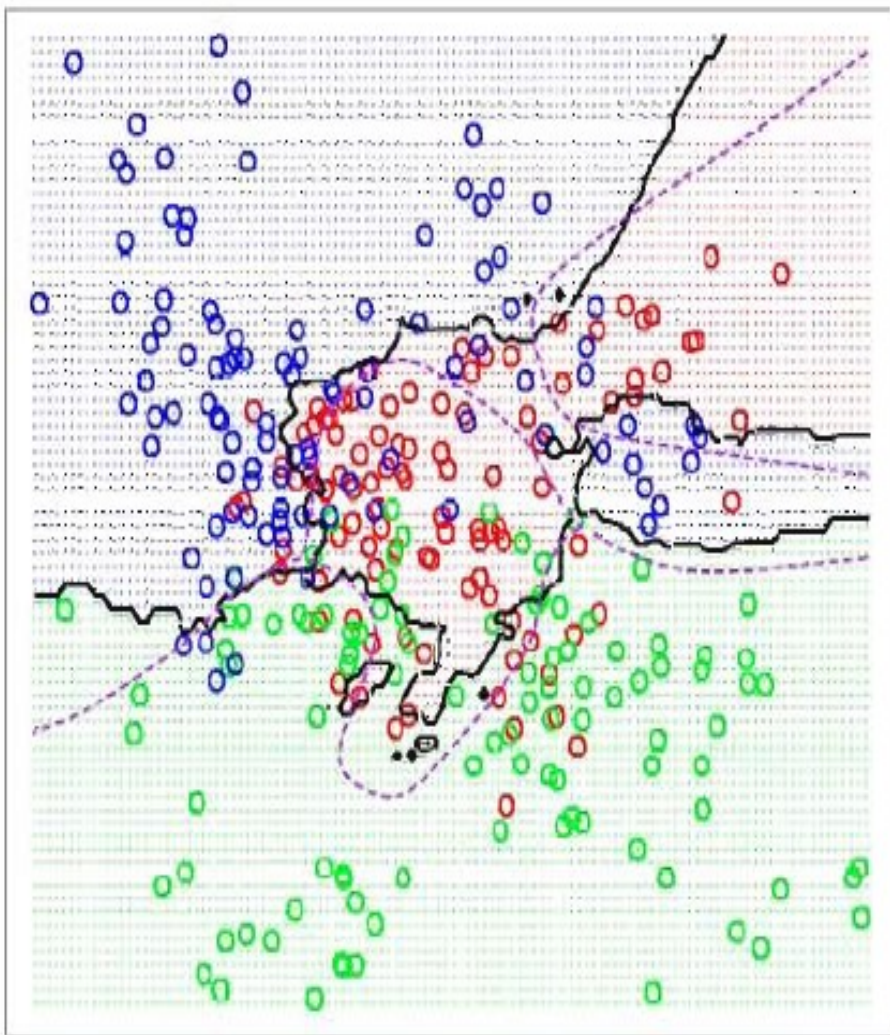
# K vecinos mas cercanos (KNN)

- Dada una nueva observación  $X_0$ , la clasifico en aquella población que posee una representación mayoritaria entre los K vecinos mas cercanos a  $X_0$ .

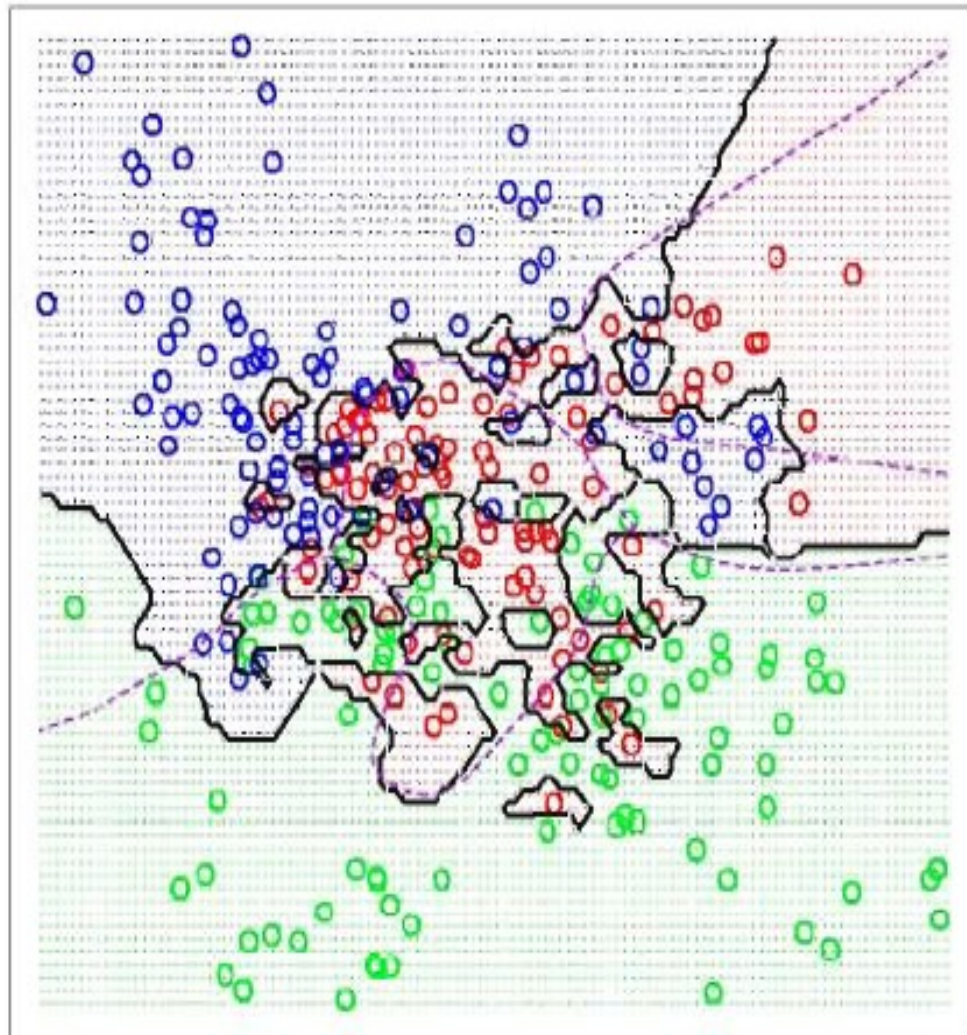


# Ejemplo gráfico

15-Nearest Neighbors



1-Nearest Neighbor



## Estimando el error de testeo

TRAIN /  
TEST

Se divide la muestra en 2 partes: muestra de entrenamiento y muestra de testeo.

Se ajusta el modelo usando la muestra de entrenamiento y el modelo ajustado se usa para predecir las respuestas de la muestra de testeo

El error cuadrático medio calculado con las observaciones de la muestra de testeo es un estimador del error de testeo.

## LOOCV

### Validación cruzada *Leave one out*

Una sola observación se usa como muestra de testeo y todas las demás como muestra de entrenamiento. Supongamos que sacamos  $(\mathbf{x}_1, y_1)$

$$ECM_1 = (y_1 - \hat{y}_1)^2$$

es un estimador "aproximadamente insesgado" del error de testeo. Este procedimiento se repite  $n$  veces, dejando afuera una observación cada vez. Obtenemos

$$ECM_1, ECM_2, \dots, ECM_n.$$

El estimador *LOOCV* del error de testeo es el promedio de estos :

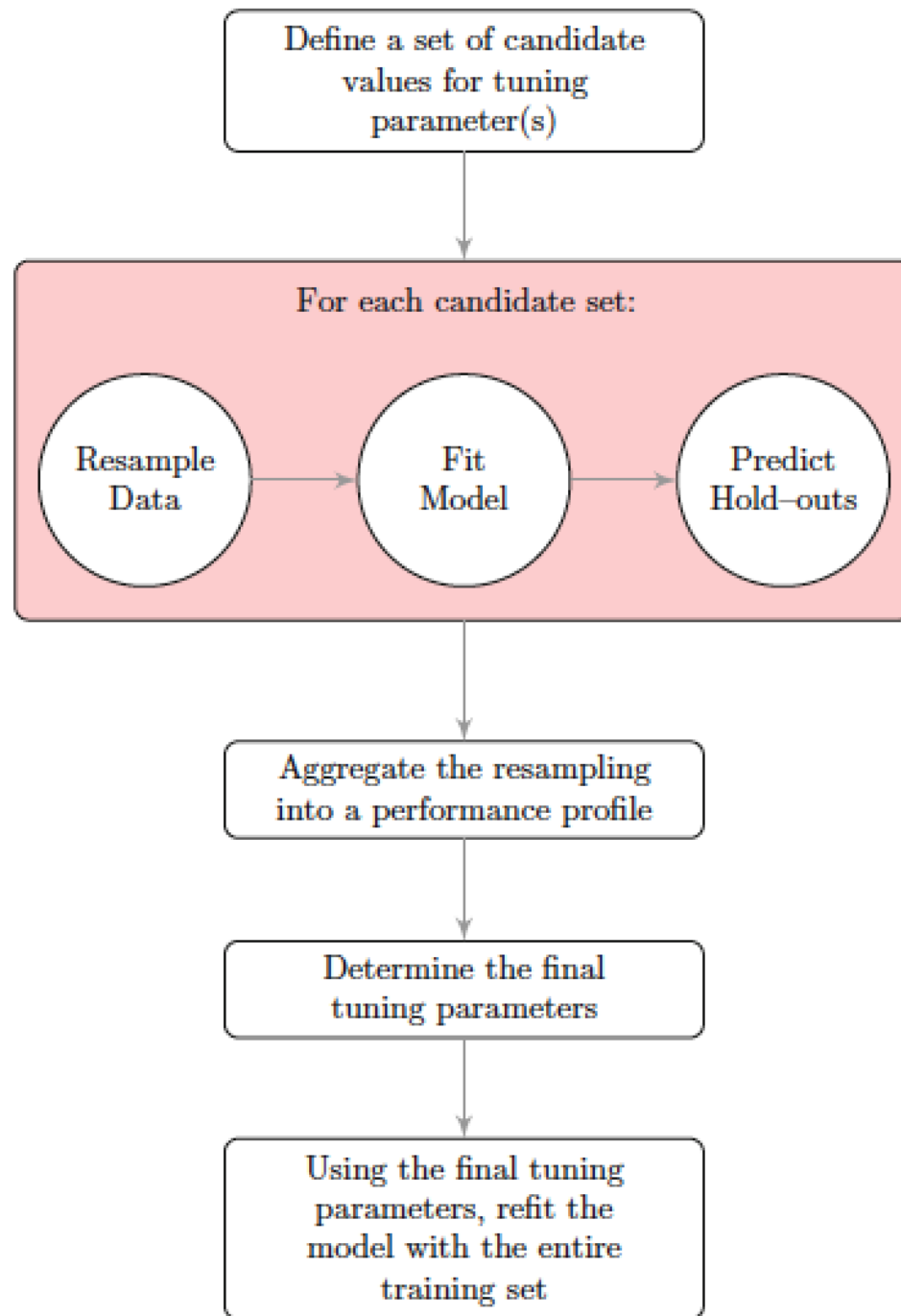
$$ECM_k = \frac{1}{n} \sum_{i=1}^n ECM_i.$$



# LOOCV Versus TRAIN/TEST

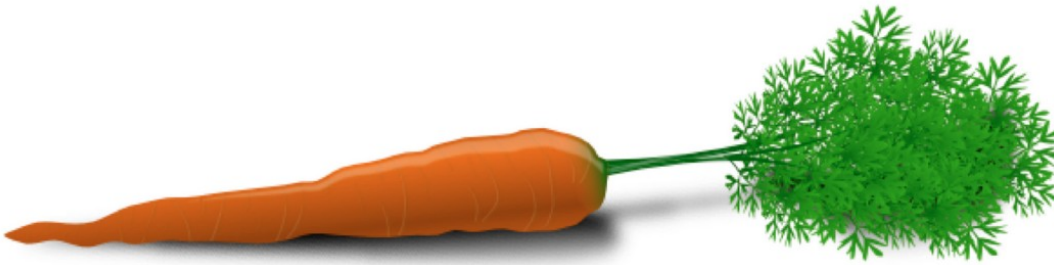
- ▶ Dividir a la muestra en submuestra de entrenamiento y de testeo da un estimador del error de testeo sesgado.
- ▶ LOOCV da estimadores aproximadamente insesgados del error de testeo, pero tienen mayor varianza.
- ▶ 5-fold o 10-fold cv es conveniente porque da un buen compromiso entre sesgo y varianza.

K-fold CV



# The caret R package

## the caret package



The **caret** package (short for Classification And REgression Training) is a set of functions that attempt to streamline the process for creating predictive models. The package contains tools for:

### Links

[train Model List](#)

### Topics

[Main Page](#)

[Data Sets](#)

[Visualizations](#)

[Pre-Processing](#)

<http://caret.r-forge.r-project.org/>



