# Descomposición en sesgo y varianza

## Aprendizaje automático

Juan David Martínez
jdmartinev@eafit.edu.co

2023



UNIVERSIDAD EAFIT

# Agenda

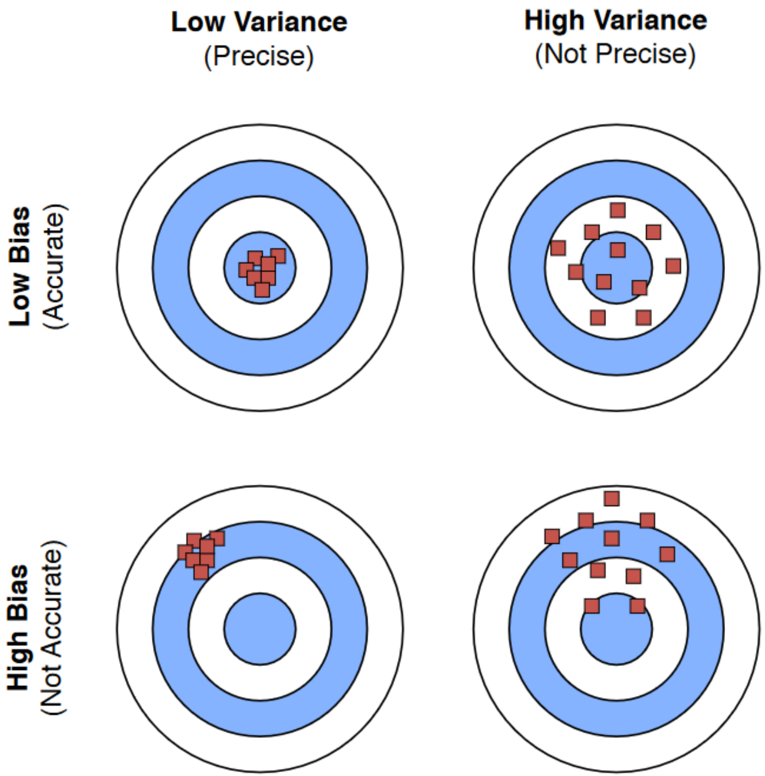
- Sobre-entrenamiento y sub-entrenamiento
- Introducción a la descomposición en sesgo y varianza
- Descomposición del error cuadrático medio
- Relación entre sesgo/varianza con sub/sobre entrenamiento
- Descomposición de sesgo y varianza de modelos de clasificación
- Otras formas de bias

# Descompisición en sesgo/varianza

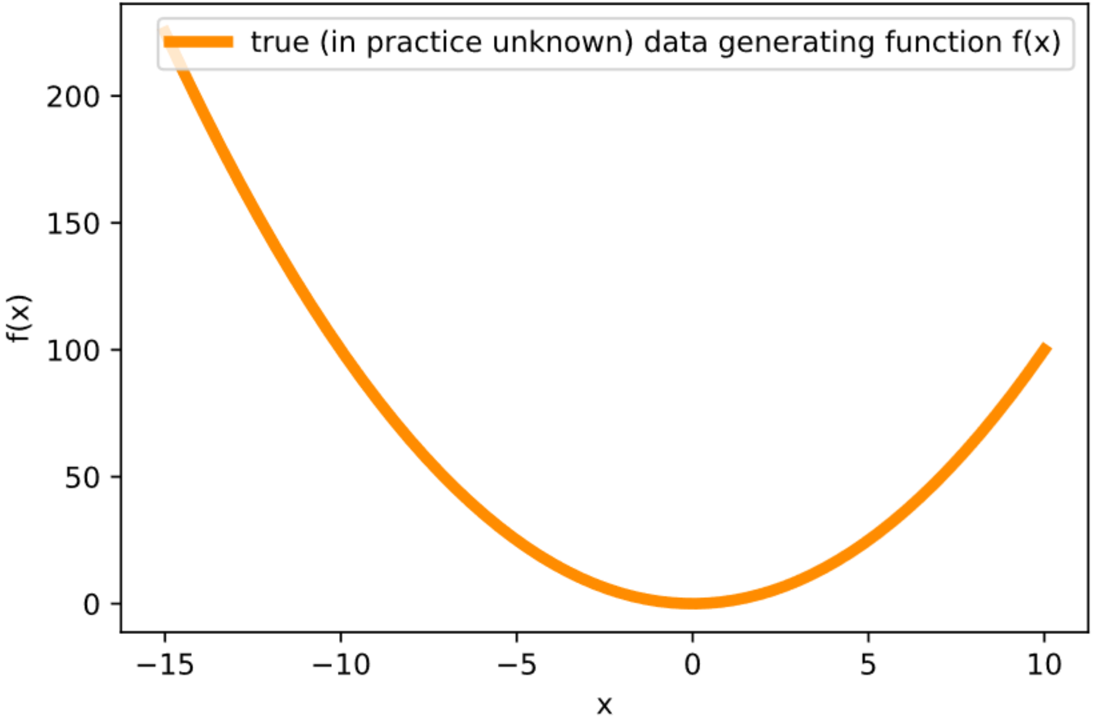
## ¿Qué significa que un modelo tenga un alto sesgo o una alta varianza?

- Descomponer la pérdida (error) en sesgo y varianza nos puede ayudar a entender mejor los algoritmos de aprendizaje, estos conceptos están relacionados con sobre y sub entrenamiento
- Ayuda a entender por qué los métodos de ensamble funcionan, en algunos casos, mejor que un solo modelo

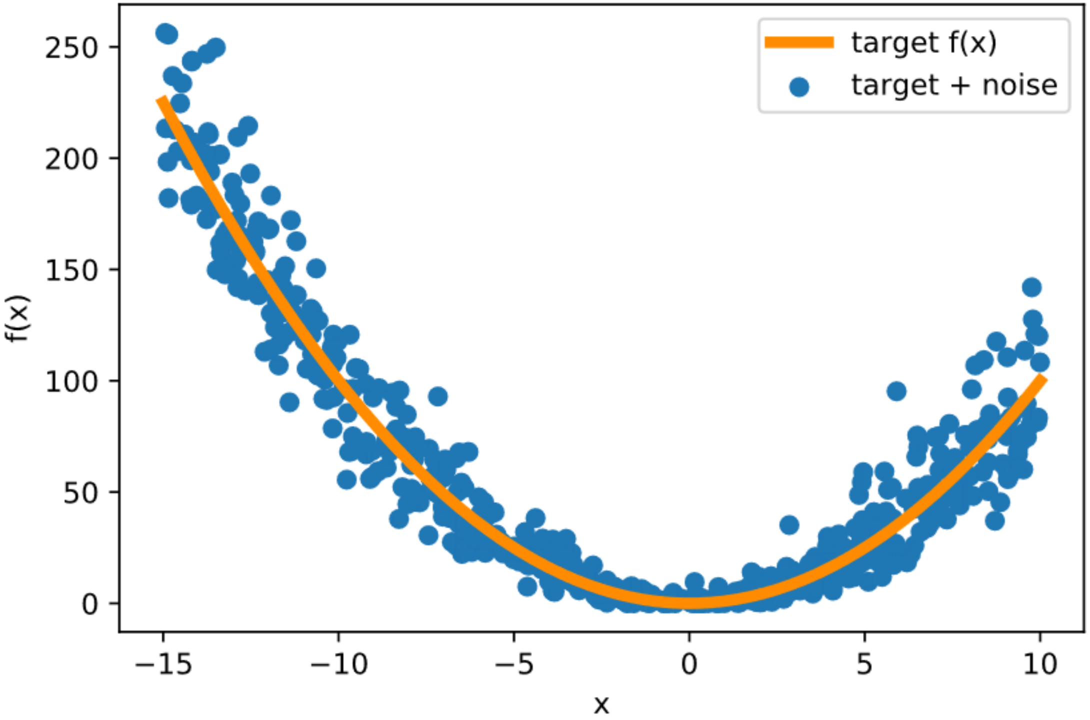
# Intución de la descomposición



# Intuición de la descomposición



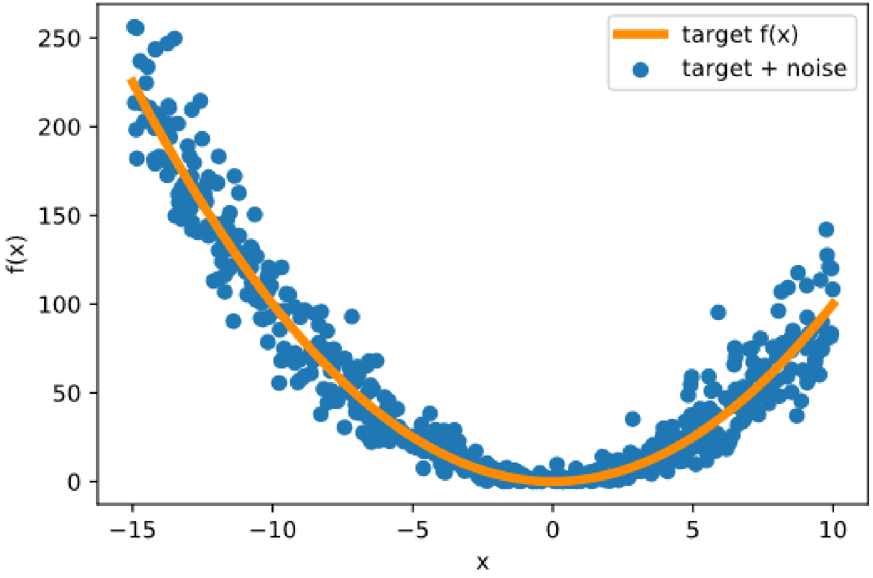
# Intuición de la descomposición

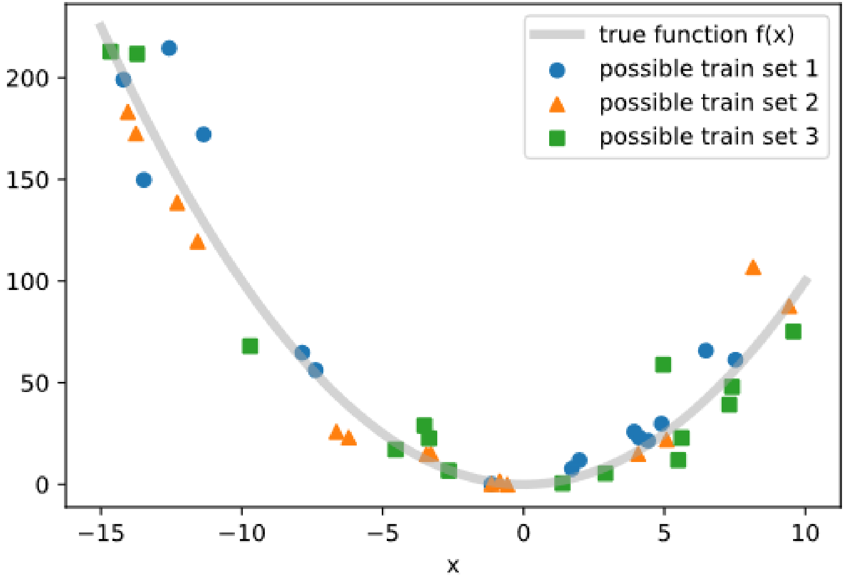




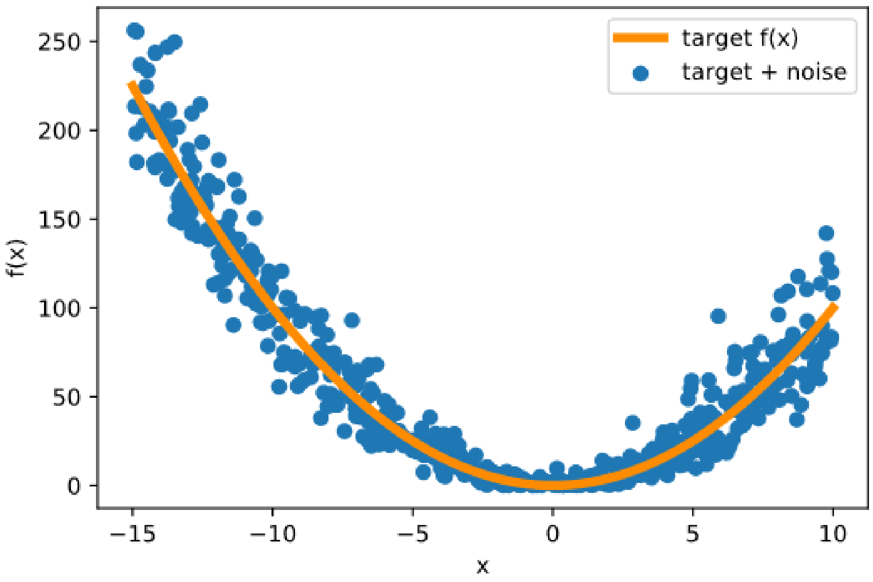

UNIVERSIDAD EAFIT

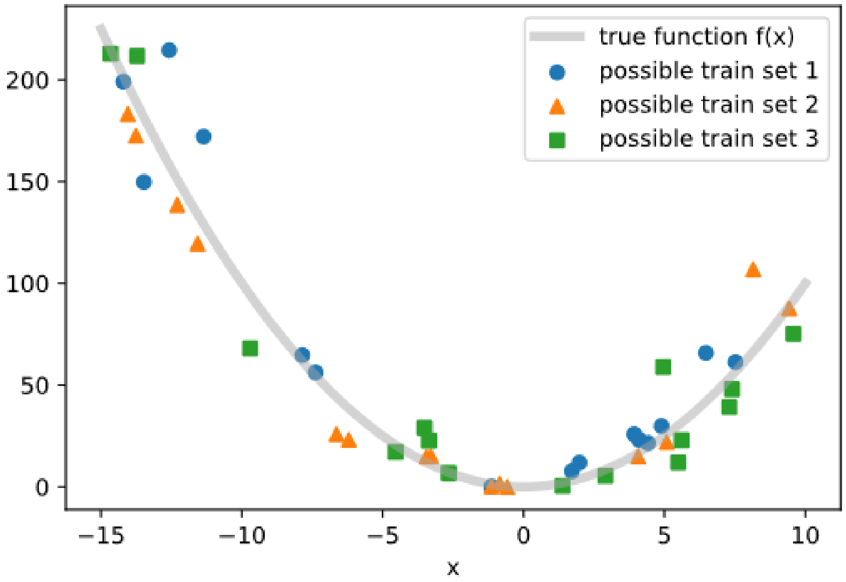
# Intuición de la descomposición



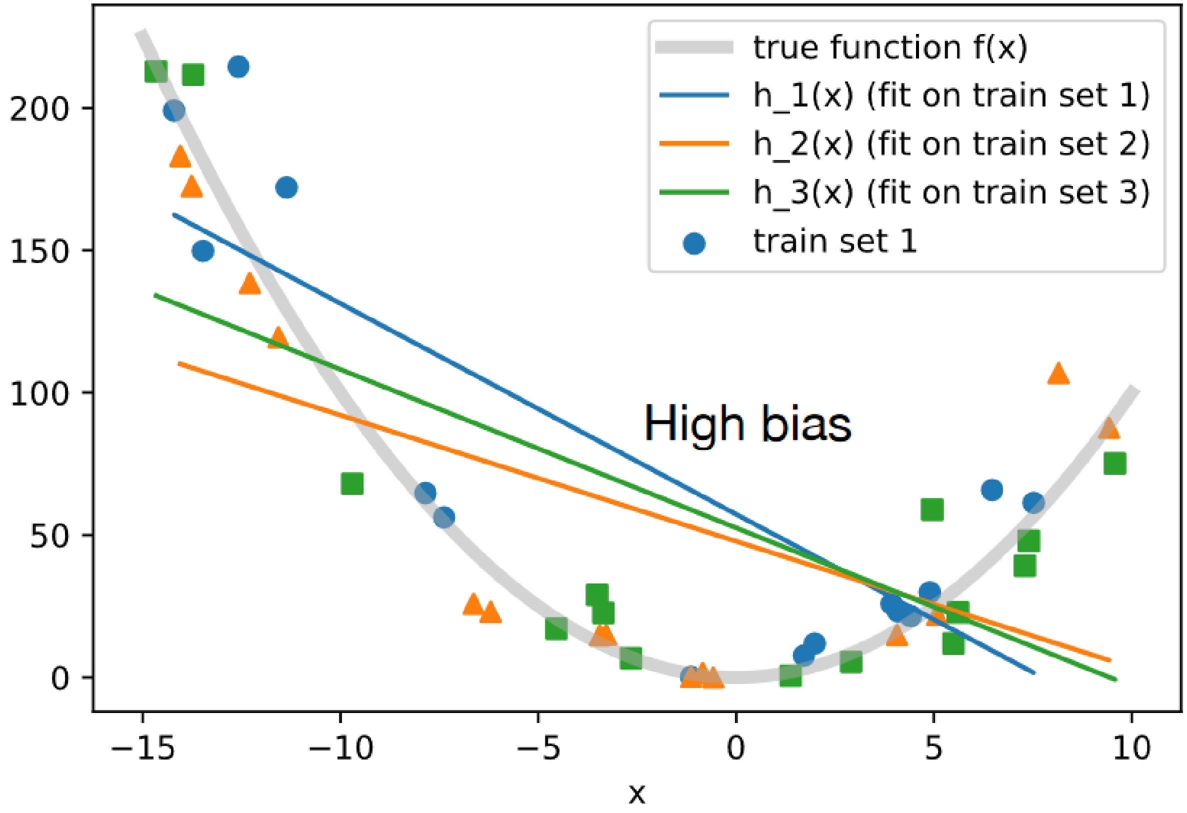



# Intuición de la descomposición



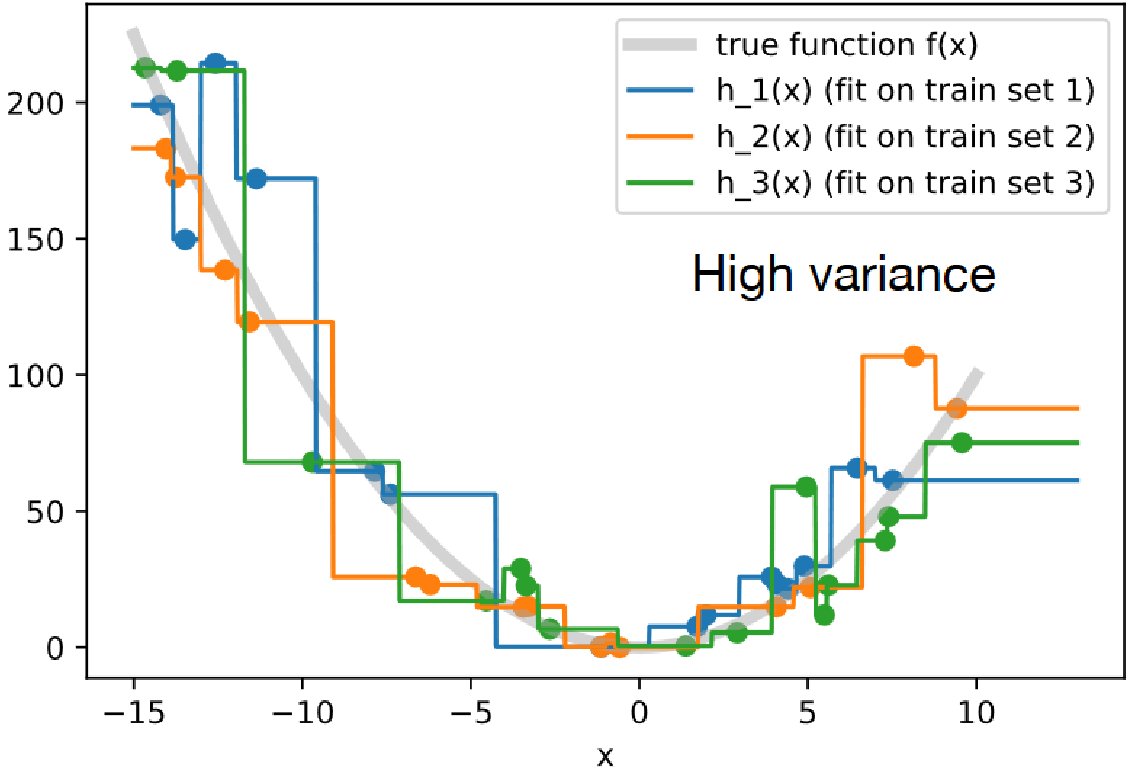



# Intuición de la descomposición

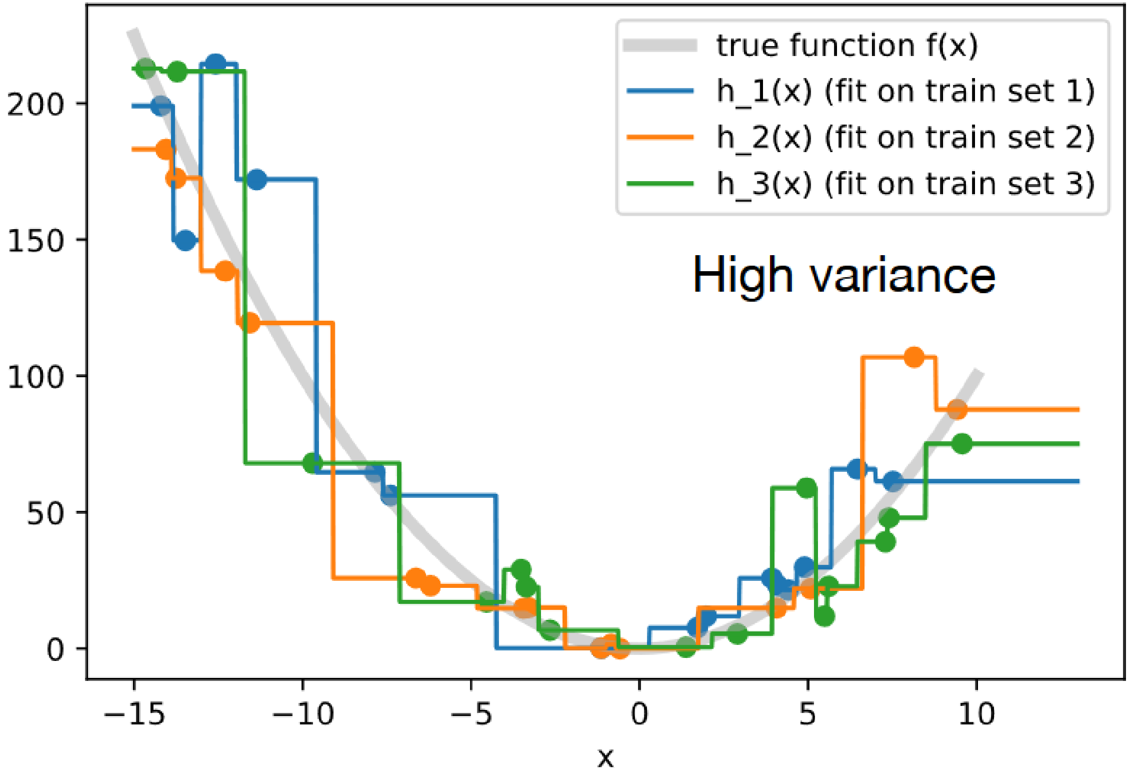



linear regression models

# Intuición de la descomposición

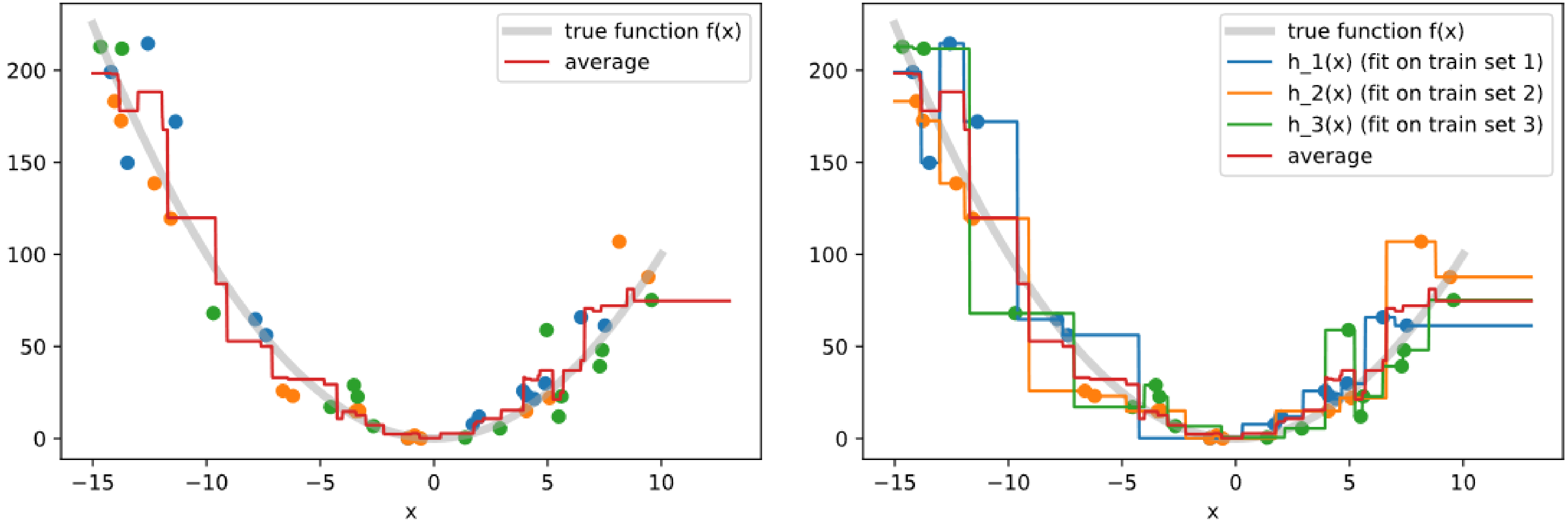


# Intuición de la descomposición




What happens if we take the average?
Does this remind you of something?

# Intuición de la descomposición



# Terminología

- $\hat{\theta}$: Estimador puntual de un parámetro
- $Bias = \mathbb{E}[\hat{\theta}] - \theta$
- $Var[\hat{\theta}] = \mathbb{E}[\hat{\theta}^2] - (\mathbb{E}[\hat{\theta}])^2$
- $Var[\hat{\theta}] = \mathbb{E}\left[(\mathbb{E}[\hat{\theta}] - \hat{\theta})^2\right]$

# Terminología

$$Bias = \mathbb{E}[\hat{\theta}] - \theta$$

$$Var[\hat{\theta}] = \mathbb{E}\left[(\mathbb{E}[\hat{\theta}] - \hat{\theta})^2\right]$$

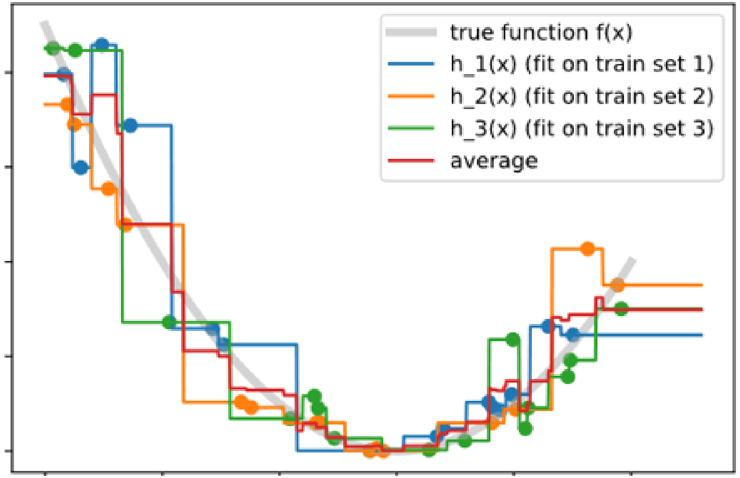
El sesgo (bias) es la diferencia entre el valor esperado del estimador obtenido de diferentes muestras de entrenamiento y el valor real. El valor esperado se calcula sobre los conjuntos de entrenamiento

La varianza da una estimación de cuanto varía la estimación a medida que variamos los datos de entrenamiento

# Descomposición en sesgo y varianza

$$error(\mathbf{x}) = bias(\mathbf{x})^2 + Var\left[\hat{f}_{\mathcal{D}}(\mathbf{x})\right] + noise(\mathbf{x})$$

$$\mathbb{E}_{\mathcal{D},\mathbf{y}}\left[\left(y - \hat{f}_{\mathcal{D}}(\mathbf{x})\right)^2\right] = \left(f(\mathbf{x}) - \bar{f}(\mathbf{x})\right)^2 + \mathbb{E}_{\mathcal{D}}\left[\left(\bar{f}(\mathbf{x}) - \hat{f}_{\mathcal{D}}(\mathbf{x})\right)^2\right] + \mathbb{E}_{\epsilon}\left[\epsilon(\mathbf{x})^2\right]$$




https://medium.com/snu-ai/the-bias-variance-trade-off-a-mathematical-view-14ff9dfe5a3c

# Relación con sub/sobre entrenamiento