

Predicción Versus Explicación

- ¿ No son lo mismo ?
- ¿ Se pueden lograr ambas simultáneamente ?
- ¿ Qué herramientas existen en ambos casos ?
 - ¿ Que podemos esperar en el futuro ?

Dr. Andrés Farall
afarall@hotmail.com

¿ Los Datos Mandan ?



EXPERT OPINION

Contact Editor: **Brian Brannon**, bbrannon@computer.org

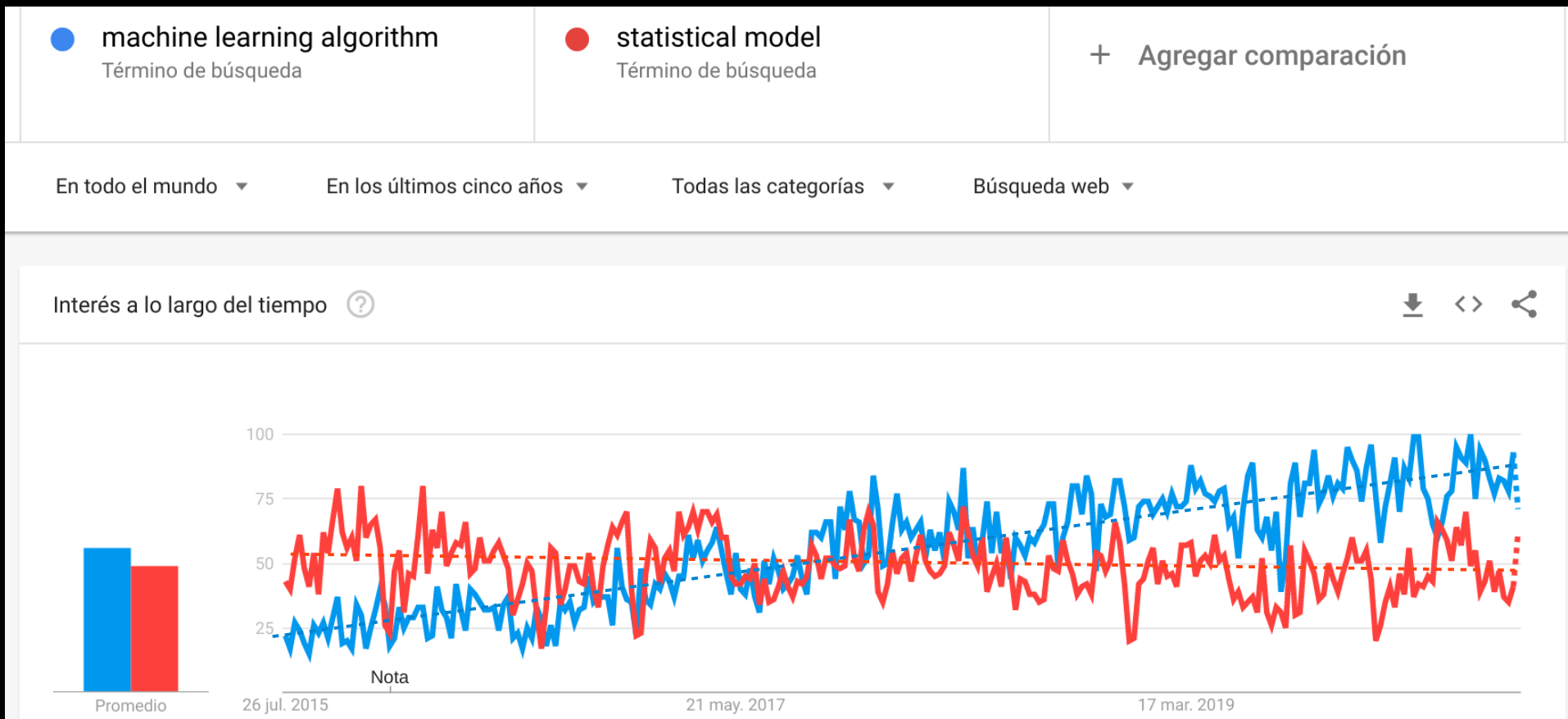
The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

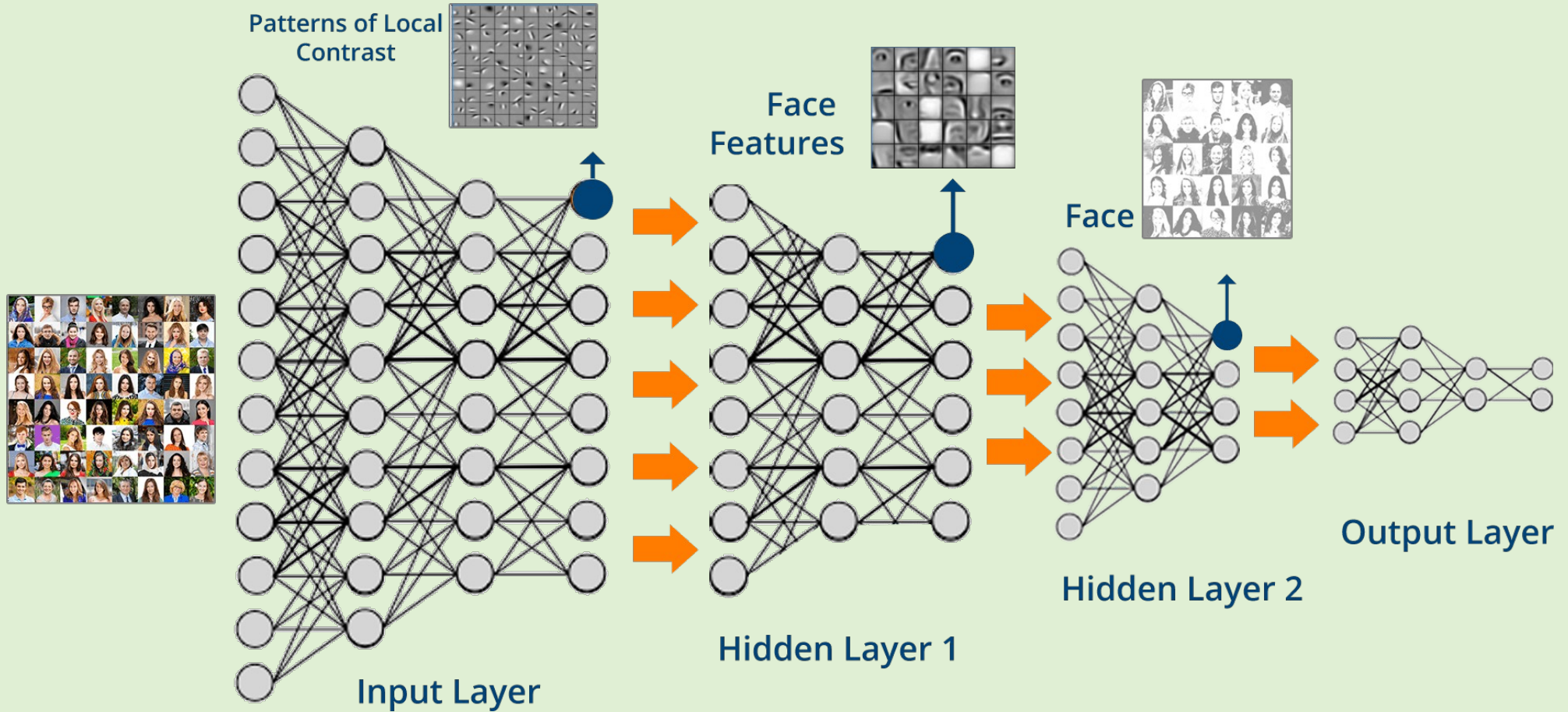
For many tasks, words and word combinations provide all the representational machinery we need to learn from text.

Google Trends

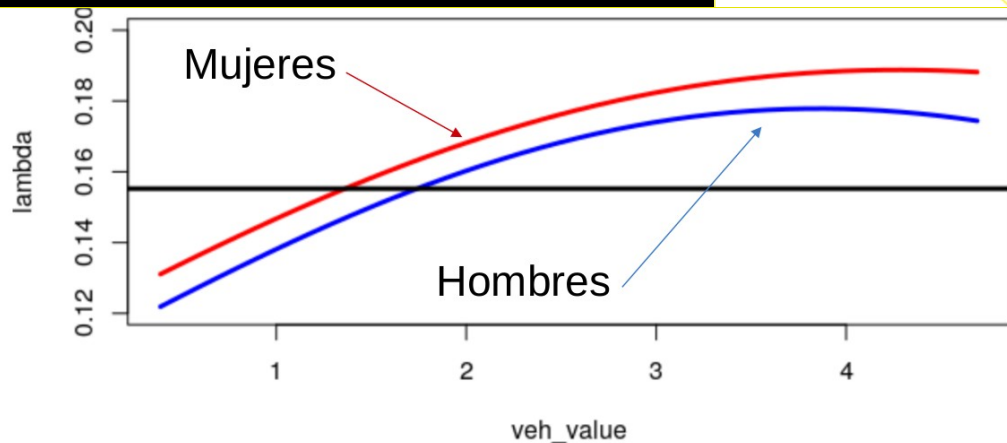
Machine Learning Vs. Modelado Estadístico



Ejemplo Paradigmático Predictivo: Reconocimiento Automático de Imágenes



Ejemplo Paradigmático Explicativo: Modelos Actuariales de Riesgo via GLM



siniestros
del riesgo i

Tasa anual específica
del riesgo i

$$\lambda_i = e^{X_i' \beta}$$

$$Y_i \sim \mathcal{P}(E_i \cdot \lambda_i)$$

Covariables
del riesgo i

La Exposición entra
como una variable mas,
con coef =1. Se lo llama

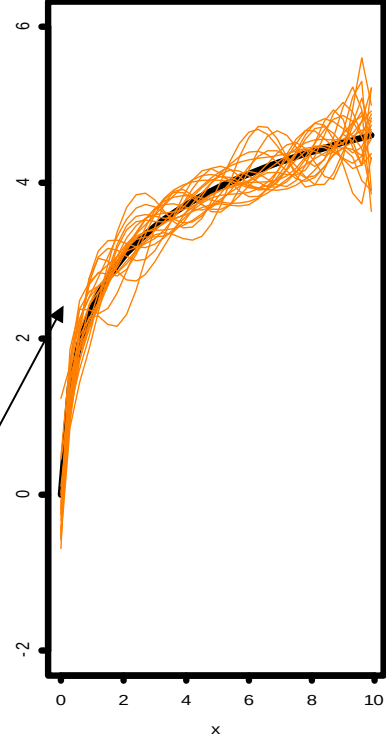
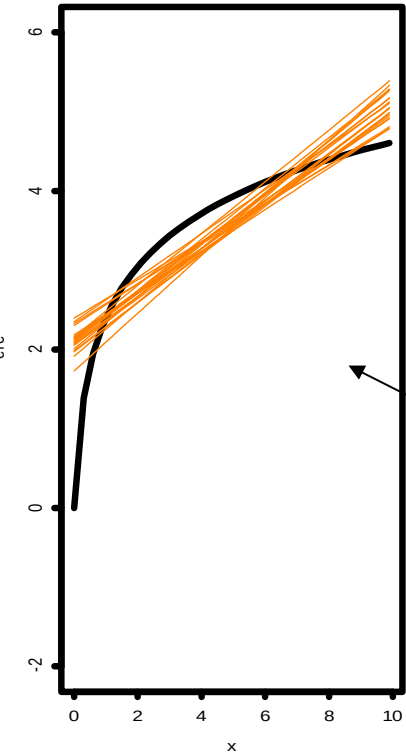
$$Y_i \sim \mathcal{P}(e^{X_i' \beta} + \log E_i)$$

OFFSET

Tradeoff Sesgo - Varianza

Mucho sesgo, poca
varianza

Poco sesgo, mucha
varianza



Mayor Sesgo

Mayor Varianza

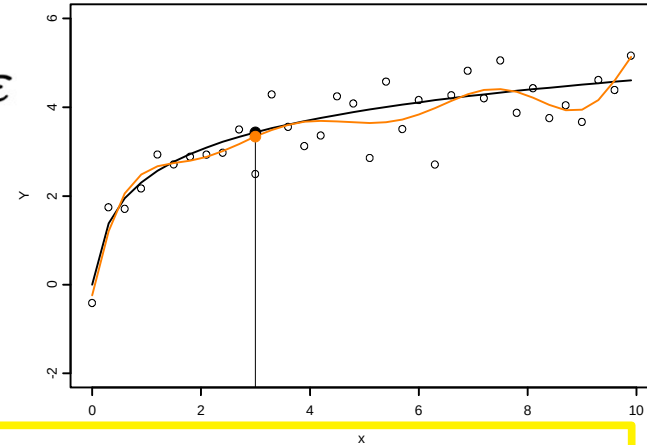
Complejidad del modelo (k)

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: $\hat{f}(x; D)$
k — Parámetro de Complejidad del Modelo



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

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

Considerando un nuevo dato fijo

$$\mathbf{E}_D [\hat{f}(x; D)] - f(x)$$

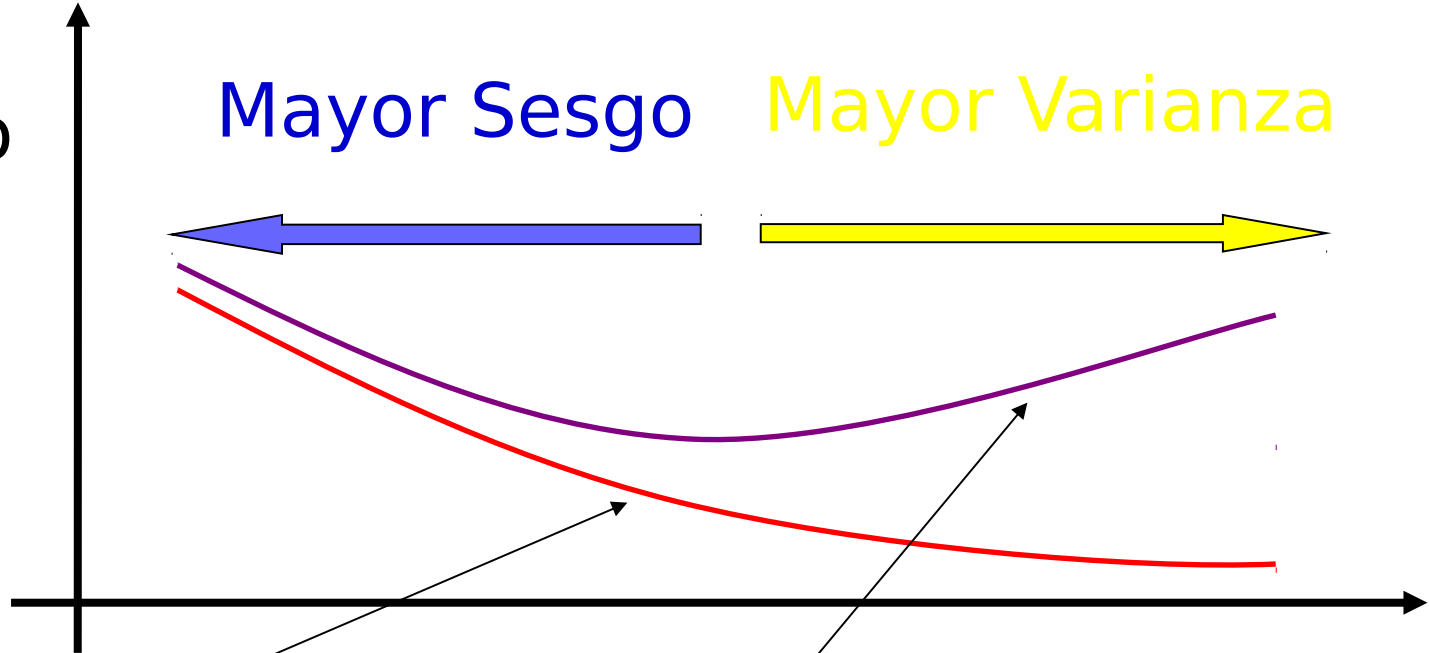
Tradeoff Sesgo - Varianza

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

Error
Cuadrático
Medio

Mayor Sesgo

Mayor Varianza



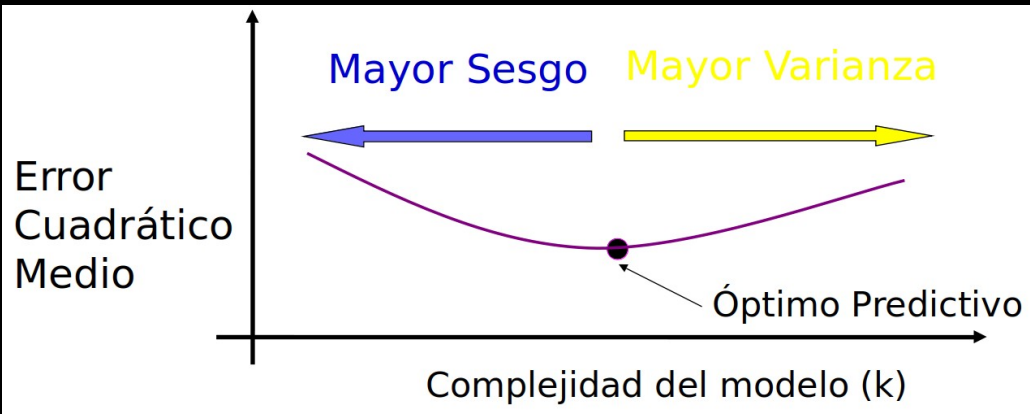
Curva muestral

Complejidad del modelo (k)

Curva verdadera

Modelado Predictivo

Intenta desarrollar mecanismos que predigan un futuro valor de Y dados ciertos valores de X .
El objetivo principal es aproximar a Y .



$$y = f(x)$$

$$Y = F(X)$$

$$\hat{Y}^* = F(X^*) \approx Y^* = f(X^*)$$

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

Modelado Explicativo

Intenta desarrollar modelos que testean, verifican e infieren relaciones causales entre variables.

El objetivo principal son las relaciones entre variables.

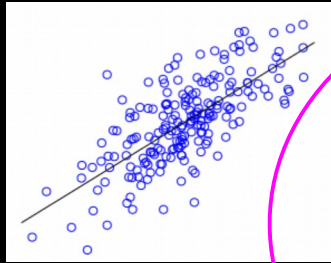


$$y = f(x)$$
$$Y = F(X)$$
$$F \approx f$$

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

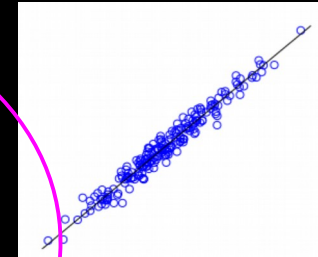
¿ Por qué Son Distintos ? “Cómo miro”

Modelo Explicativo



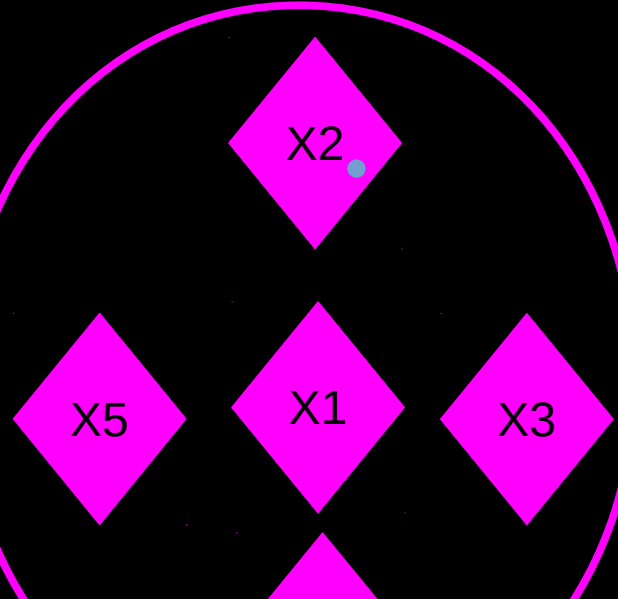
$$y = f(x)$$
$$Y = F(X)$$
$$F \approx f$$

Modelo Predictivo



$$y = f(x)$$
$$Y = F(X)$$
$$\widehat{Y}^* = F(X^*) \approx Y^* = f(X^*)$$

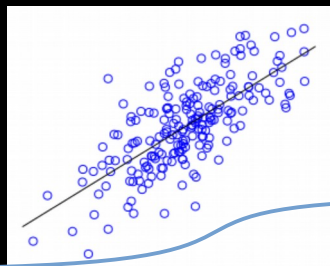
Y



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

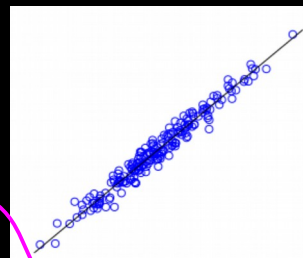
¿ Por qué Son Distintos ? “Qué miro”

Modelo Explicativo



Y

Modelo Predictivo



X

Z2

Z5

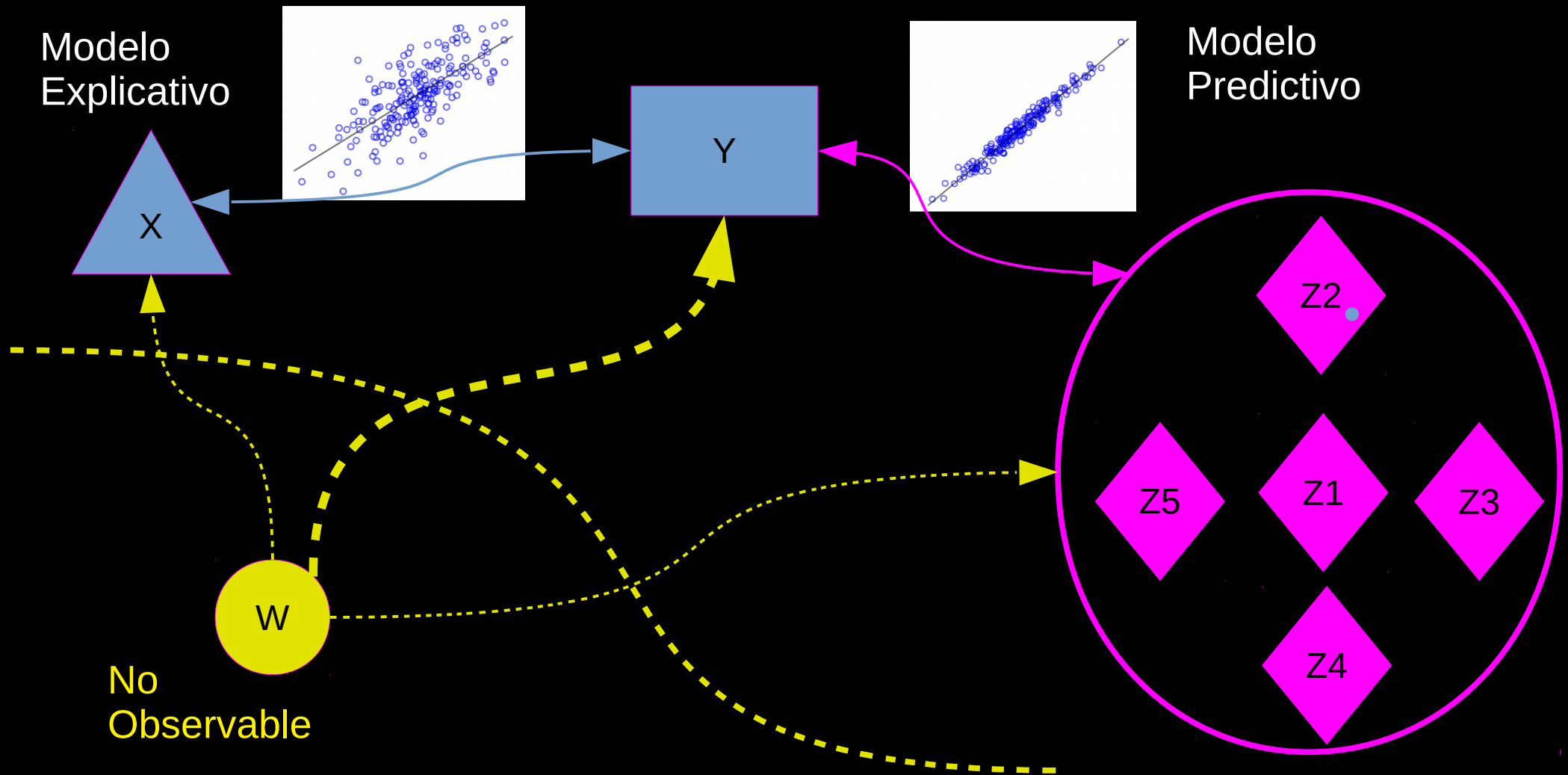
Z1

Z3

Z4

W

No Observable



¿ Qué Hacer Ahora ?

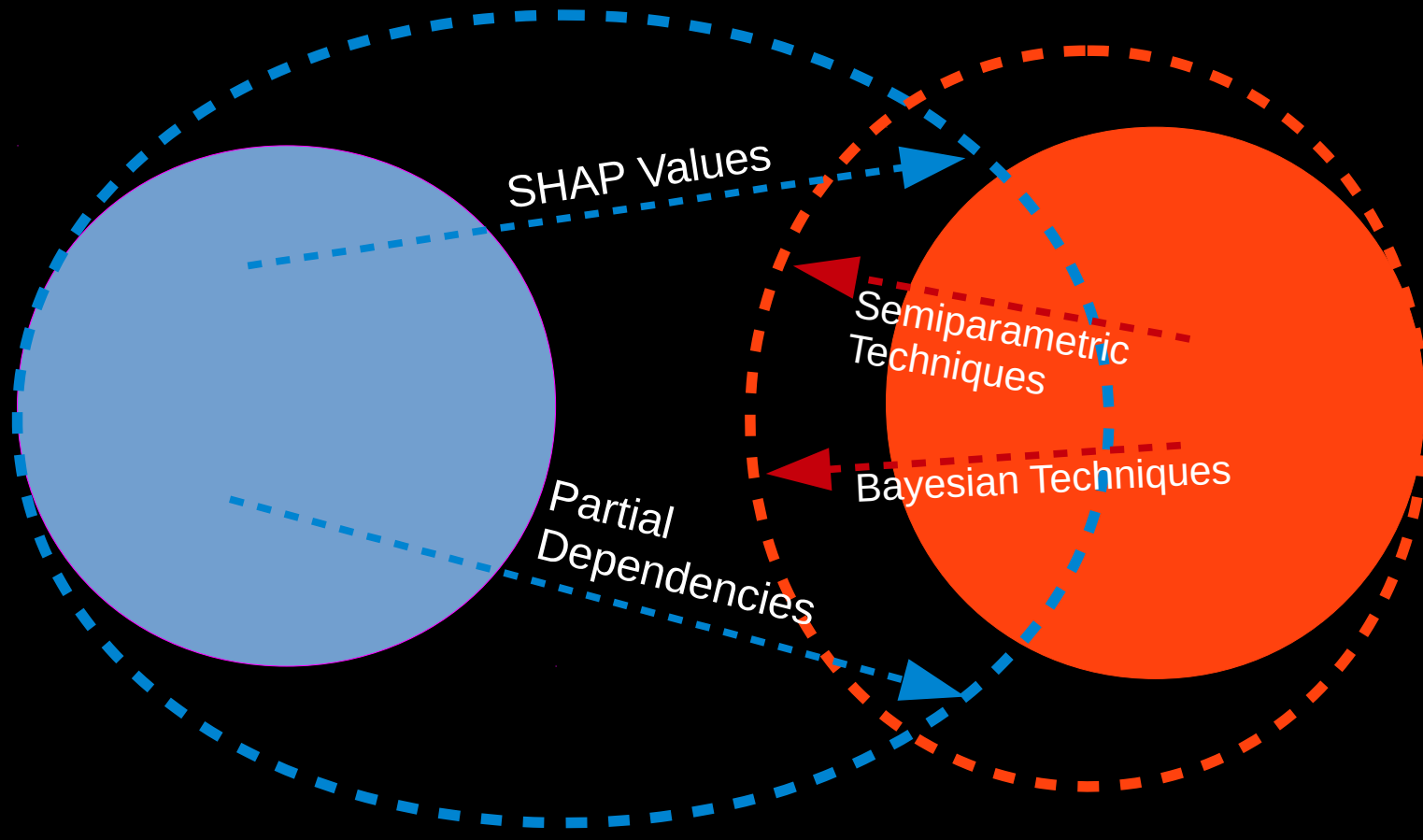
Machine Learning Vs. Modelado Estadístico

K-Vécinos mas Cercanos (KNN)
Redes Neuronales (ANN)
Random Forest (RF)
Support Vector Machines (SVM)
Gradient Boosting (XGB)

Modelos Lineales (LM)
Modelos Lineales Generalizados (GLM)
Modelos Aditivos Generalizados (GAM)
Árboles de Decisión (CART)
Modelos Bayesianos
Modelos Semiparamétricos
Sistemas de Ecuaciones Estructurales (SEM)

¿ Qué Podemos Esperar en el Futuro ?

Machine Learning Vs. Modelado Estadístico



Referencias

Shmueli, G. (2010). To explain or to predict?. *Statistical science*, 25(3), 289-310.

Halevy, A., Norvig, P., & Pereira, F. (2009).
The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2), 8-12.

Predicting vs. Explaining
And Why Data Science Needs More “Half-Bayesians” ([https://
towardsdatascience.com/predicting-vs-explaining-
69b516f90796](https://towardsdatascience.com/predicting-vs-explaining-69b516f90796))

Gracias !