

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Estadística y computación para metagenómica

Victor Muñiz Sánchez

victor\_m@cimat.mx

Centro de Investigación en Matemáticas.  
Unidad Monterrey.

## Generalidades

### Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Sobre ésta parte del curso...

# Temario

Victor Muñiz

## Generalidades

### Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## ① Introducción y conceptos generales

- Machine learning (ML) supervisado y no supervisado
- Elementos de teoría de decisión estadística
- Métricas de desempeño para ML supervisado
- El compromiso entre precisión e interpretabilidad
- El compromiso entre sesgo y varianza

## ② Modelos de clasificación

- ① El modelo lineal y sus limitaciones
- ② Regresión logística
- ③ Redes neuronales
- ④ Hiperplanos separadores óptimos y Máquinas de Soporte Vectorial
- ⑤ Selección de modelos y regularización
- ⑥ Modelos aditivos y métodos relacionados
  - Arboles de decisión
  - Bagging y random forests

# Generalidades

## Generalidades

### Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

- En esta parte del curso mostraremos algunos métodos y conceptos básicos de aprendizaje máquina (ML: machine learning) y reconocimiento estadístico de patrones para el análisis de datos multivariados en general, y con aplicaciones en datos metagenómicos en particular.
- El enfoque será práctico, revisando los conceptos importantes de los modelos y poniendo énfasis en el elemento computacional.
- Los pre-requisitos: conocimientos básicos de modelos estadísticos (inferencia y regresión), álgebra lineal, cálculo de varias variables, conocimientos de programación.

# Generalidades

Victor Muñiz

## Generalidades

### Introducción

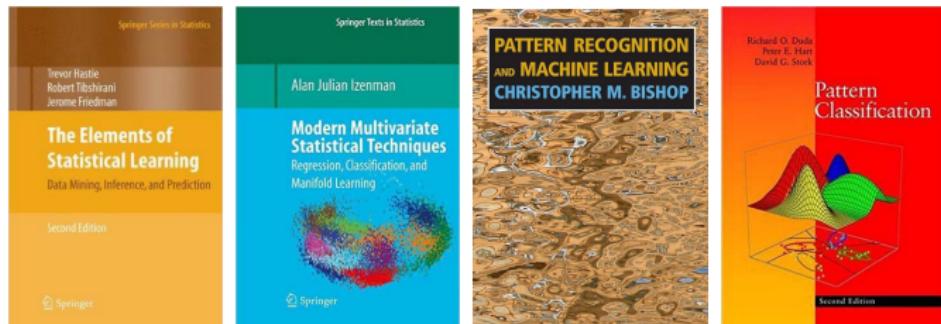
Elementos de teoría de decisión estadística

Métricas de desempeño

Prediction-interpretability tradeoff

Bias-variance tradeoff

## Libros: los principales...



# Generalidades

## Generalidades

### Introducción

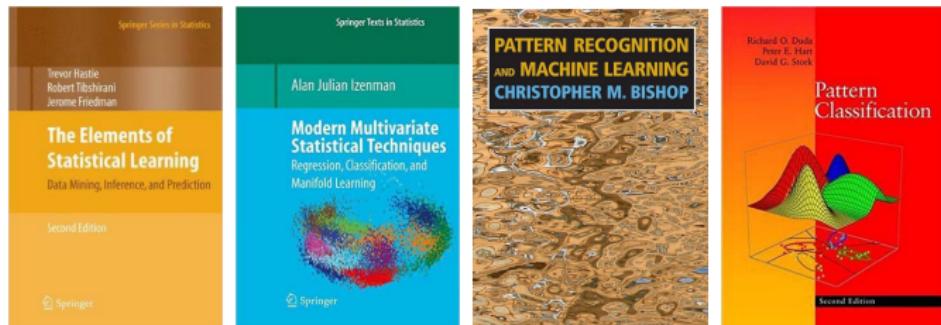
Elementos de teoría de decisión estadística

Métricas de desempeño

Prediction-interpretability tradeoff

Bias-variance tradeoff

## Libros: los principales...



Usaremos algunos artículos para conceptos más técnicos y también videos donde se podrán ilustrar detalles de algunos temas que no se cubran (por cuestiones de tiempo) en ésta parte del curso.

# Generalidades

## Software:



### Generalidades

#### Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Generalidades

## Software:



Python y la infraestructura para cómputo científico y ciencia de datos



# Generalidades

## Generalidades

### Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Instalación local.

- La opción más óptima: instalar python  
<https://www.python.org/> y usar el editor de tu preferencia (vi, emacs, pycharm, spyder, jupyter-notebook, etc).
- La opción más rápida (recomendada para iniciar), instalar la suite Anaconda:  
<https://www.anaconda.com/>
- En cualquier caso, recomiendo ampliamente crear **virtual environments** para el curso y/o proyectos específicos que requieran a su vez, librerías específicas.

Instalar librerías: En Anaconda, se puede hacer directamente en el framework. También puedes hacerlo en consola con los comandos pip y conda (si tienes Anaconda).

# Generalidades

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



- Colaboratory permite escribir y ejecutar código de Python en un navegador
- Sin configuración requerida
- Acceso gratuito a GPU y TPU
- Facilidad para compartir

## Generalidades

### Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Introducción y conceptos generales

# Conceptos básicos I

Victor Muñiz

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Machine learning (ML): modelos que aprenden de los datos

Algunos conceptos relacionados:

# Conceptos básicos II

Victor Muñiz

Generalidades

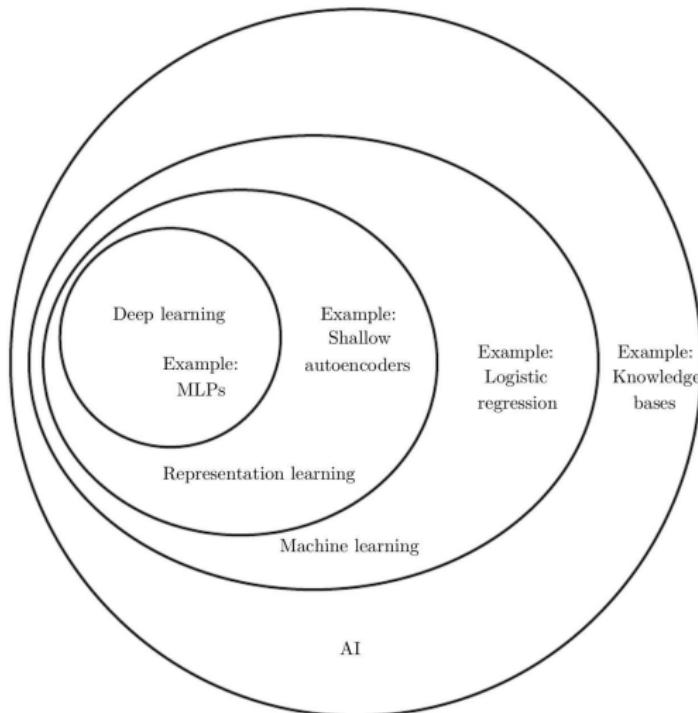
## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

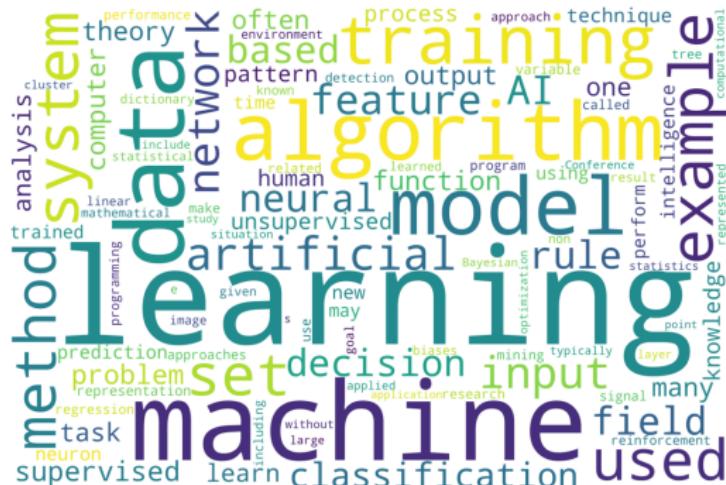


Goodfellow et al., Deep learning, 2016.

## Conceptos básicos III

## Introducción

- Machine learning

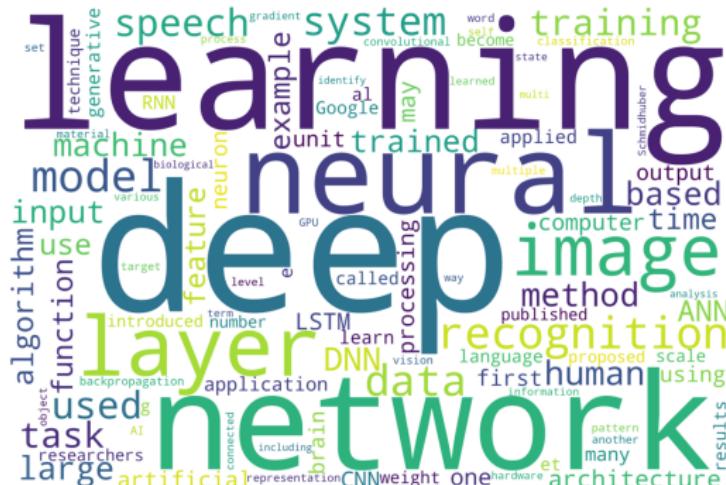


Fuente: wikipedia.

## Conceptos básicos IV

## Introducción

- Deep learning



Fuente: wikipedia.

## Conceptos básicos V

## Introducción

- Representation learning



Fuente: wikipedia.

# Conceptos básicos VI

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## ¿Qué aprenden éstos modelos?

- A predecir un valor de respuesta
- A reconocer patrones existentes
- A generar nuevos datos
- Una representación de los datos

# Conceptos básicos VII

## Generalidades

### Introducción

Elementos de teoría de  
decisión estadística

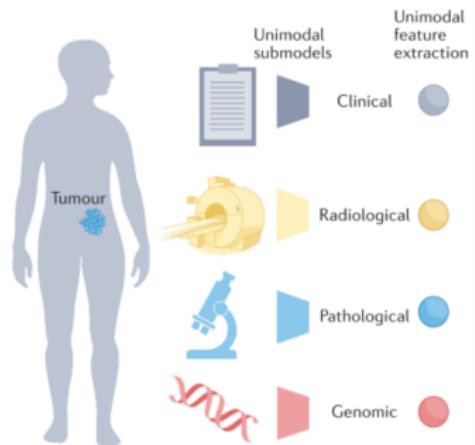
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

### Por ejemplo:

- ❖ Predicción: ¿estás enfermo de un tumor? [si o no]
- ❖ Está en función de:
  - ❖ Tus datos de salud clínicos
  - ❖ Tus datos de estudios radiológicos, de laboratorio y genómicos



# Conceptos básicos VIII

Victor Muñiz

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

- ❖ Predicción: ¿cuál es la siguiente palabra?  
[buena, aburrida, equis]
- ❖ Está en función de:
  - ❖ La secuencia de texto que has escrito
  - ❖ Tu historial y “estilo” de escritura



# Conceptos básicos IX

Victor Muñiz

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

- ❖ ¿Cuántos grupos de imágenes encuentras?
- ❖ Está en función de:
  - ❖ Los pixeles de las imágenes
  - ❖ Las “formas” que existen



# Conceptos básicos X

Generalidades

Introducción

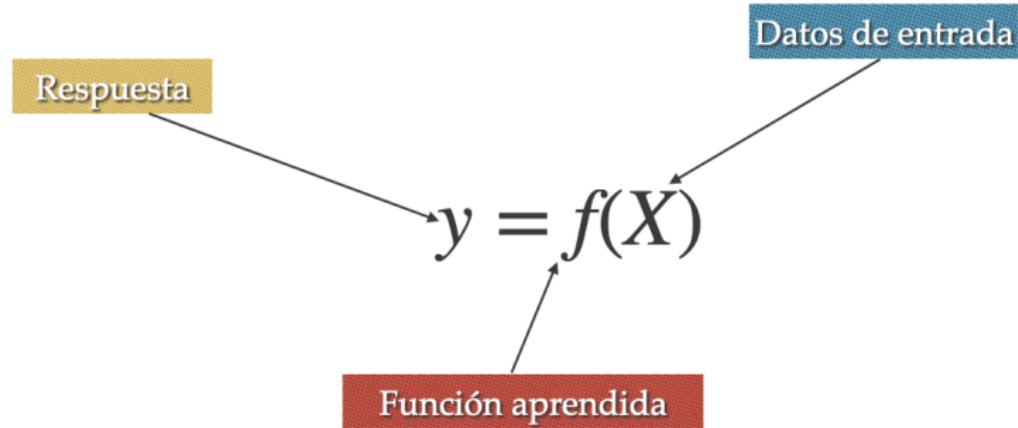
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Matemáticamente, todo lo anterior lo expresamos como



# Conceptos básicos XI

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

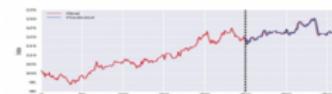
Prediction-interpretability  
tradeoff

Bias-variance tradeoff

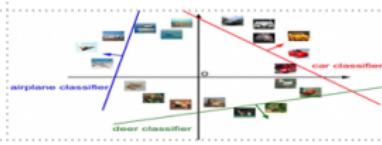
## Variable de respuesta

$y$

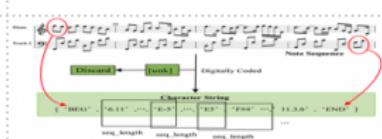
Puede ser un valor *continuo*  
(Ej: pronóstico en series de  
tiempo)



Puede ser un valor *categórico*  
(Ej: clasificación de imágenes)



Puede ser un *objeto*  
(Ej: una secuencia de música)



# Conceptos básicos XII

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

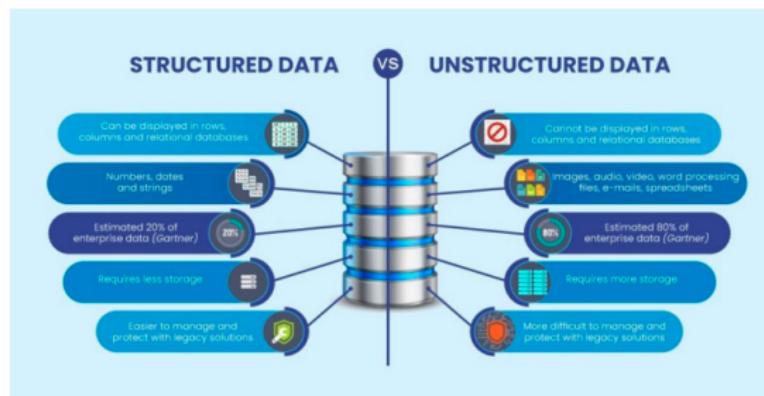
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Datos de entrada X

Pueden ser estructurados o no-estructurados



# Conceptos básicos XIII

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

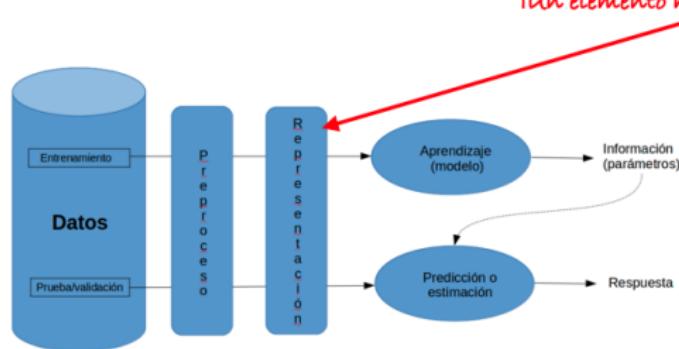
Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Función aprendida $f(X)$

Se aprende mediante un *modelo* probabilístico  
del fenómeno de interés

iun elemento muy importante!



# Conceptos básicos XIV

## El modelo

Un modelo es una descripción abstracta de un sistema [o fenómeno] concreto. Un modelo matemático utiliza conceptos y lenguaje matemático.

- Determinístico: cada conjunto de estados de las variables está determinado de manera única. Siempre produce la misma respuesta para un conjunto de condiciones iniciales dadas.
- Probabilístico: existen factores aleatorios, por lo tanto, las variables no están determinadas de manera única. La respuesta del modelo se obtiene mediante una distribución de probabilidad.

Dentro de los modelos probabilísticos o estadísticos, generalmente hay dos tipos:

# Conceptos básicos XV

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

- Paramétricos. Son modelos (learner) que resume, obtiene información o modela datos a través de una colección de parámetros para obtener  $f$ . Los modelos paramétricos suponen un conjunto finito de parámetros  $\theta$ . Esto significa que el modelo ya conoce la cantidad de parámetros que requiere, independientemente de sus datos. Dados los parámetros, la inferencia sobre  $x$  es independiente de los datos observados  $D$ :

$$P(x|\theta, D) = P(x|\theta),$$

entonces,  $\theta$  captura todo lo que debe conocerse sobre los datos.

# Conceptos básicos XVI

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

- No paramétricos. Suponen que  $f$  (por ejemplo, la distribución de los datos) no se puede definir en términos de un conjunto finito de parámetros especificados a priori, sino que se determina a partir de los datos.  
El número y la naturaleza de los parámetros son flexibles y no están fijados de antemano. A menudo se pueden definir asumiendo un  $\theta$  de dimensión infinita. Generalmente pensamos en  $\theta$  como **una función**. La cantidad de información que  $\theta$  puede capturar sobre los datos  $D$  puede crecer a medida que crece la cantidad de datos. Esto los hace más flexibles.

# Conceptos básicos XVII

Victor Muñiz

Generalidades

Introducción

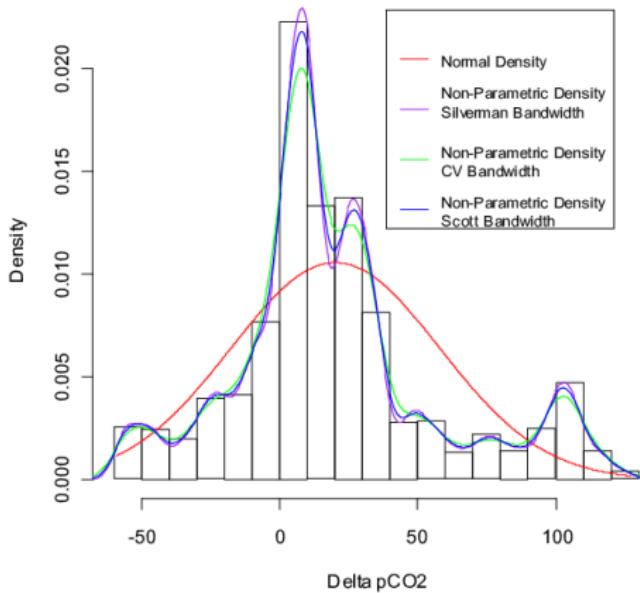
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Fit of Parametric and Non-Parametric Density Estimation



# Conceptos básicos XVIII

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## ¿Qué aprende el modelo?

Básicamente, aprende **similaridades** o diferencias entre los datos.

Ejemplo: Palmer Penguins



| species | island    | bill_length_mm | bill_depth_mm | flipper_length_mm | body_mass_g | sex    |        |
|---------|-----------|----------------|---------------|-------------------|-------------|--------|--------|
| Adelie  | Torgersen | 39.1           | 18.7          | 181.0             | 3750.0      | male   |        |
| Adelie  | Torgersen | 39.5           | 17.4          | 186.0             | 3800.0      | female |        |
| Adelie  | Torgersen | 40.3           | 18.0          | 195.0             | 3250.0      | female |        |
| Adelie  | Torgersen | 36.7           | 19.3          | 193.0             | 3450.0      | female |        |
| Adelie  | Torgersen | 39.3           | 20.6          | 190.0             | 3650.0      | male   |        |
| ...     | ...       | ...            | ...           | ...               | ...         | ...    |        |
| 339     | Chinstrap | Dream          | 55.8          | 19.8              | 207.0       | 4000.0 | male   |
| 340     | Chinstrap | Dream          | 43.5          | 18.1              | 202.0       | 3400.0 | female |
| 341     | Chinstrap | Dream          | 49.6          | 18.2              | 193.0       | 3775.0 | male   |
| 342     | Chinstrap | Dream          | 50.8          | 19.0              | 210.0       | 4100.0 | male   |
| 343     | Chinstrap | Dream          | 50.2          | 18.7              | 198.0       | 3775.0 | female |

333 rows × 7 columns

## Generalidades

### Introducción

Elementos de teoría de  
decisión estadística

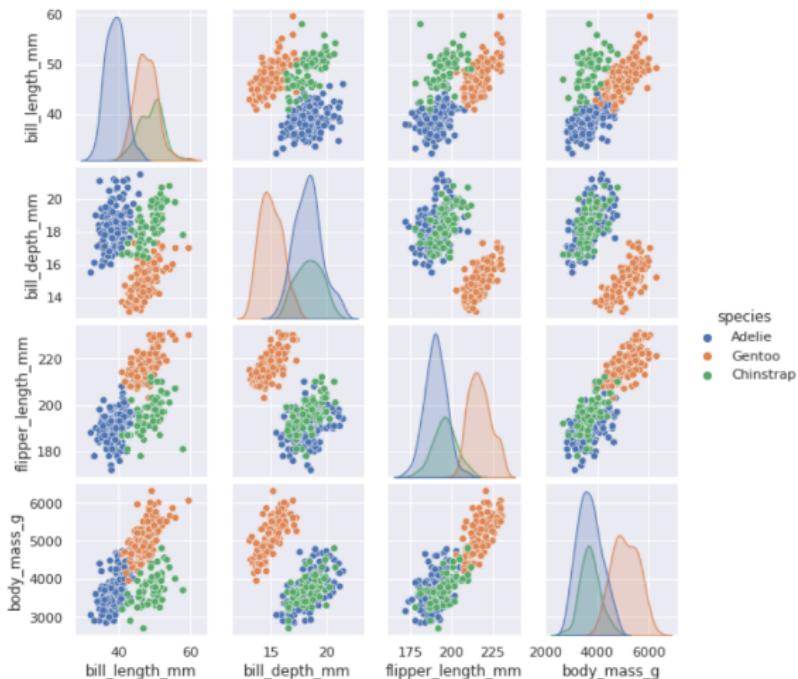
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Conceptos básicos XIX

Aprendizaje no supervisado: detectar patrones en los datos.



# Conceptos básicos XX

Generalidades

Introducción

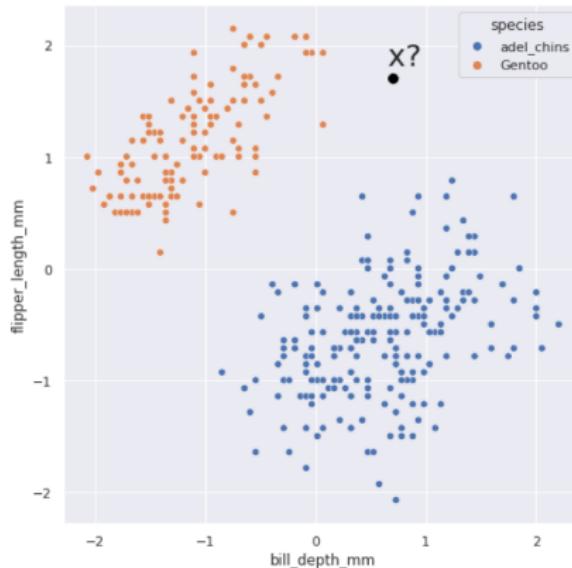
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

**Aprendizaje supervisado:** predecir datos nuevos.



# Conceptos básicos XXI

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

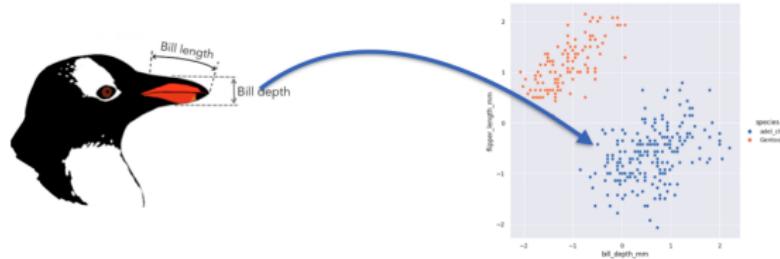
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

**Requisito:** los datos deben ser capaces de **representarse** como objetos matemáticos.

**La representación de los datos** es un elemento primordial. Para datos estructurados, es relativamente fácil obtener una representación matemática (vectorial) de los mismos.



# Conceptos básicos XXII

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Aunque hay un factor importante que hay que considerar con datos en alta dimensión.

Considera por ejemplo, los datos de OUTs del Metagenomic Geolocation Challenge CAMDA 2020.

Ubicación de las muestras



# Conceptos básicos XXIII

Victor Muñiz

## Generalidades

### Introducción

Elementos de teoría de decisión estadística

Métricas de desempeño

Prediction-interpreability tradeoff

Bias-variance tradeoff

| #                      | Name_Labels                        | Latitude  | Longitude   | X4479792 | X4471655 | X9160 | X4468191 | X188079 | X43648   | X310311  | X216610 | X1515145 | X43487   | X818401  | X1762785 | X6160    | X154208 | X314566  |
|------------------------|------------------------------------|-----------|-------------|----------|----------|-------|----------|---------|----------|----------|---------|----------|----------|----------|----------|----------|---------|----------|
| 1011aken2              | Ward's<br>Sudley's<br>Hollingshead | 23.184800 | -117.349100 | 0.030308 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.004762 | 0.071721 | 0.009080 | 0.000000 | 0.0     | 0.001273 |
| 1011aken2              | Ward's<br>Sudley's<br>Hollingshead | 23.184800 | -117.349100 | 0.020000 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.020392 | 0.023872 | 0.001980 | 0.000000 | 0.0     | 0.001718 |
| 1011aken1              | Ward's<br>Sudley's<br>Hollingshead | 23.184800 | -117.349100 | 0.010104 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.004563 | 0.004610 | 0.000300 | 0.000000 | 0.0     | 0.00194  |
| 1011aken3              | Ward's<br>Sudley's<br>Hollingshead | 23.184800 | -117.349100 | 0.009308 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.004945 | 0.004945 | 0.000000 | 0.000000 | 0.0     | 0.001243 |
| 1011aken2              | Ward's<br>Sudley's<br>Hollingshead | 23.184800 | -117.349100 | 0.010102 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.004512 | 0.004607 | 0.000000 | 0.000000 | 0.0     | 0.001945 |
| 1011aken1              | Ward's<br>Sudley's<br>Hollingshead | 23.184800 | -117.349100 | 0.009302 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.004684 | 0.004684 | 0.000000 | 0.000000 | 0.0     | 0.001908 |
| 1011.2011.1205         | cluster que no tiene ningún label  | 68.081033 | -140.822223 | 0.000000 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0     | 0.000008 |
| 1011.2011.1205         | cluster que no tiene ningún label  | 68.080933 | -140.822223 | 0.000000 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0     | 0.000008 |
| 1011.2011.1205         | cluster que no tiene ningún label  | 68.080833 | -140.822223 | 0.000000 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0     | 0.000008 |
| 1011.2011.1205         | cluster que no tiene ningún label  | 68.080833 | -140.822223 | 0.000000 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0     | 0.000008 |
| 1011.2011.1205         | cluster que no tiene ningún label  | 68.080833 | -140.822223 | 0.000000 | 0.000000 | 0.0   | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.0     | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0     | 0.000008 |
| 3853 rows x 19 columns |                                    |           |             |          |          |       |          |         |          |          |         |          |          |          |          |          |         |          |

¿Porqué es complicado trabajar con datos en alta dimensión?

Y en general, ¿porqué se complica la aplicación de los métodos de aprendizaje cuando pasamos a datos con mayores dimensiones?

# Conceptos básicos XXIV

Victor Muñiz

Generalidades

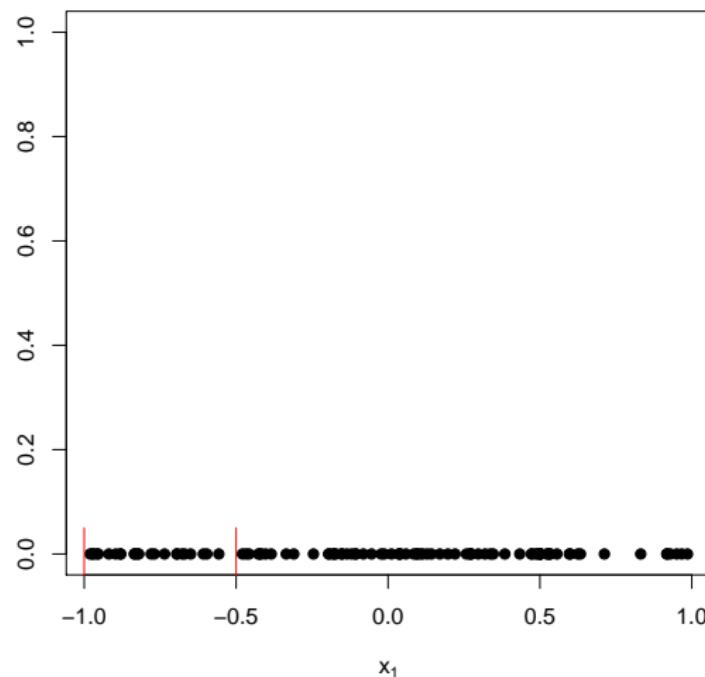
## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Conceptos básicos XXV

Victor Muñiz

Generalidades

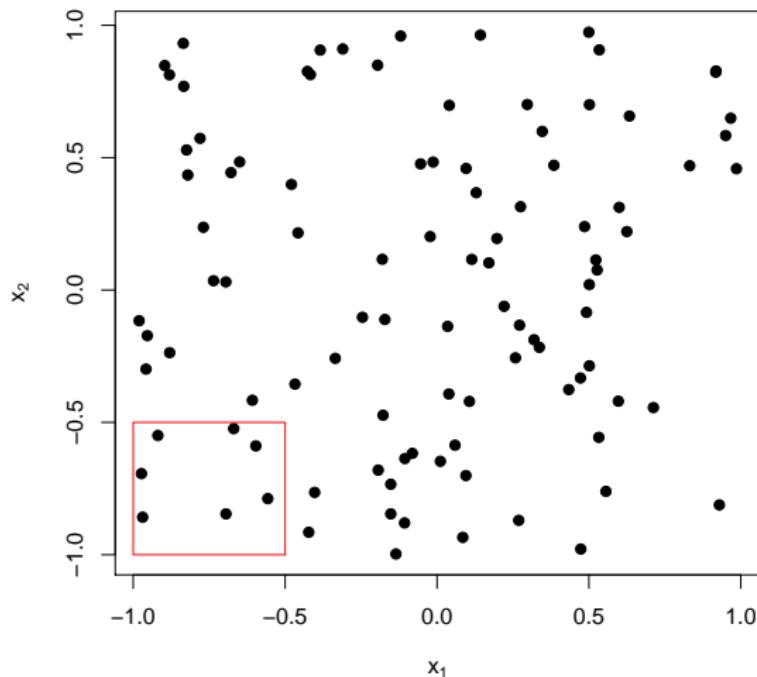
## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Conceptos básicos XXVI

Victor Muñiz

Generalidades

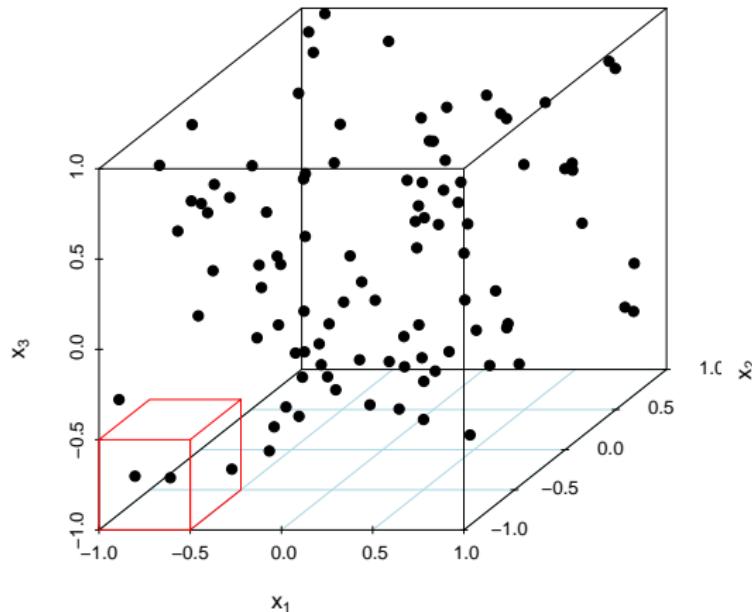
## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Conceptos básicos XXVII

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## La maldición de la dimensionalidad

- Nunca tendremos suficientes datos para cubrir cada parte de un espacio de entrada de alta dimensión, lo que hace extremadamente difícil determinar qué parte del espacio es importante para definir alguna relación entre los datos y cuál no lo es.
- Equivalentemente, la longitud necesaria en cada variable para cubrir cierto volumen de los datos, aumenta conforme aumenta la dimensión de los mismos.
- Es decir, en altas dimensiones, los datos tienden a ser “ralos” (sparse).

# Conceptos básicos XXVIII

Considera datos uniformemente distribuidos en un hipercubo con longitud  $A$  en cada lado.

Si queremos abarcar una proporción  $p$  del hipercubo con un subcubo de longitud  $A - \epsilon$  en cada lado, vemos que:

- La longitud proporcional de cada lado del subcubo es

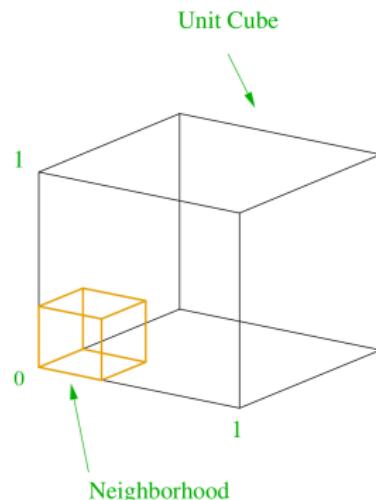
$$A_d(p) = p^{1/d} \rightarrow 1 \text{ cuando } d \rightarrow \infty,$$

donde  $d$  es la dimensión de los datos.

Y de forma equivalente:

- La proporción del volúmen entre los dos cubos es

$$\frac{A^d - (A-\epsilon)^d}{A^d} = 1 - \left(1 - \frac{\epsilon}{A}\right)^d \rightarrow 1 \text{ cuando } d \rightarrow \infty.$$



Ejemplo para  $A = 1$  (hipercubo unitario) y diferente dimensionalidad.

Supongamos que queremos capturar el 10 % de los datos...

# Conceptos básicos XXIX

Victor Muñiz

Generalidades

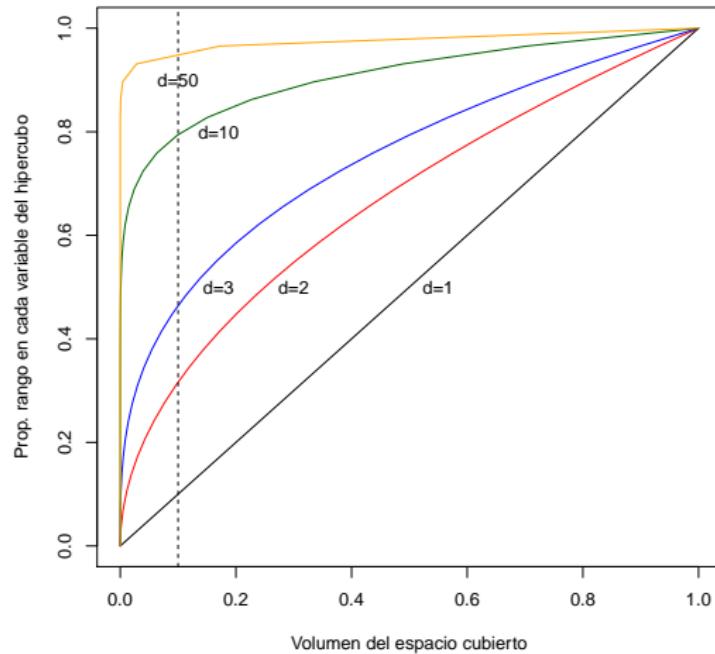
## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Conceptos básicos XXX

Victor Muñiz

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Ejemplo para  $A = 1$  (hipercubo unitario) y diferente dimensionalidad.

Supongamos que queremos capturar el 10 % de los datos...

# Conceptos básicos XXXI

Victor Muñiz

Generalidades

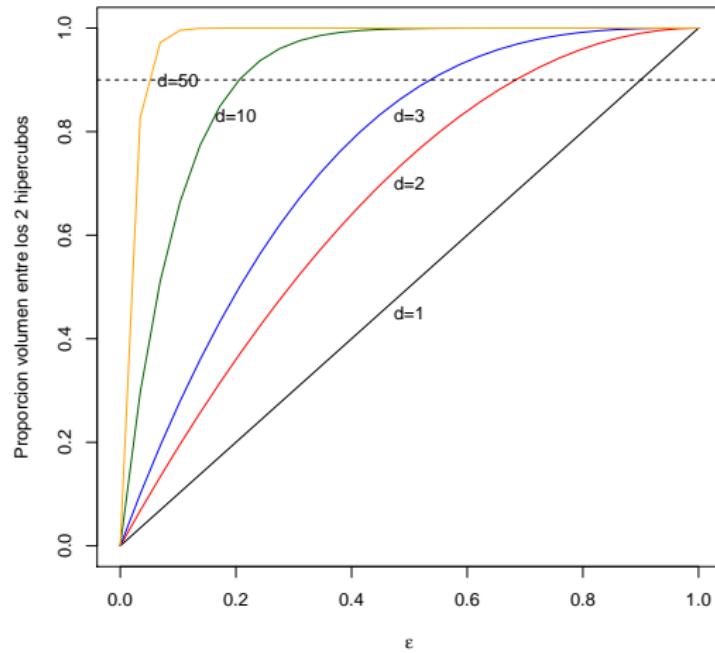
Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



Entonces:

# Conceptos básicos XXXII

Victor Muñiz

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

- Conforme aumenta el número de dimensiones, casi todo el volumen dentro de una subregión del hipercubo definido por el espacio de entrada se aproxima a la superficie del hipercubo, en vez del centroide u otra región.
- El mismo fenómeno ocurre cuando consideramos regiones esféricas.

# Conceptos básicos XXXIII

Generalidades

Introducción

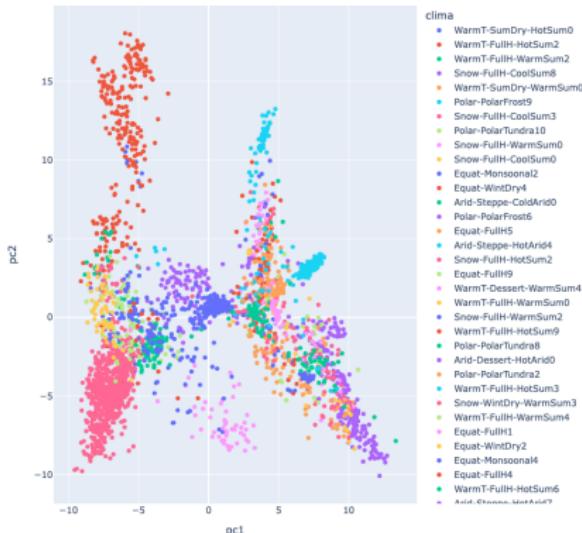
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Algunas representaciones en baja dimensión para los datos de OTUs



# Conceptos básicos XXXIV

Victor Muñiz

## Generalidades

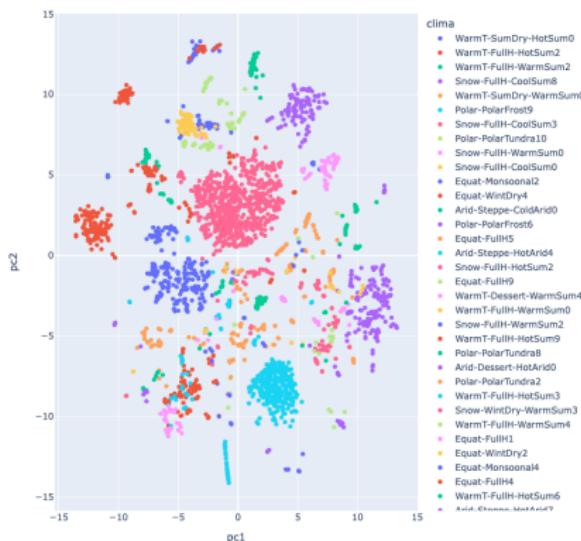
### Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Conceptos básicos XXXV

Victor Muñiz

Generalidades

## Introducción

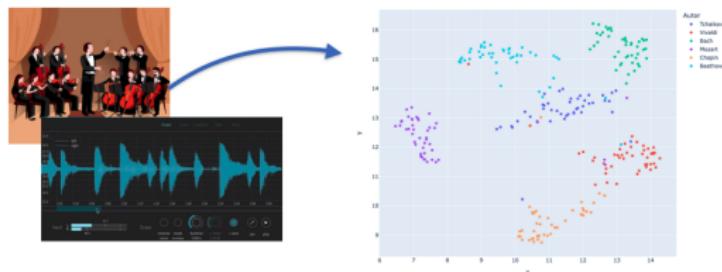
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Para datos no estructurados, la representación puede ser  
todo un reto:



# Conceptos básicos XXXVI

Victor Muñiz

## Generalidades

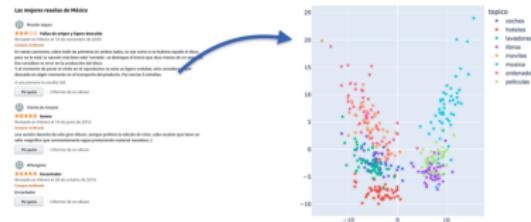
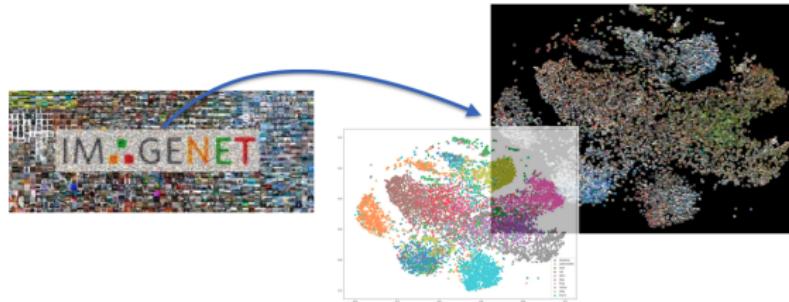
### Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Conceptos básicos XXXVII

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

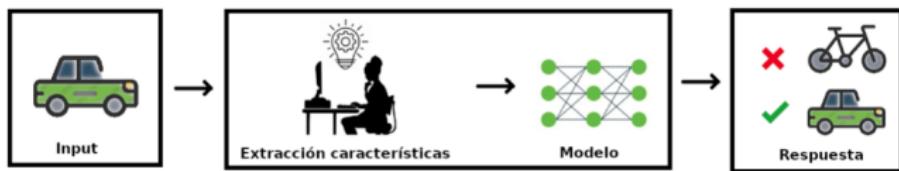
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Aprendizaje de representaciones

Este concepto ha sido fundamental para el desarrollo de los métodos de aprendizaje actuales



# Conceptos básicos XXXVIII

Victor Muñiz

Generalidades

## Introducción

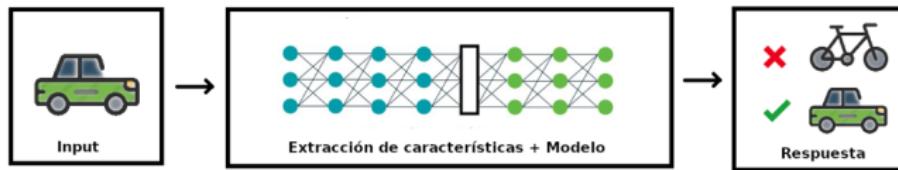
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Este concepto ha sido fundamental para el desarrollo de los métodos de  
aprendizaje actuales



# Conceptos básicos XXXIX

Victor Muñiz

Generalidades

## Introducción

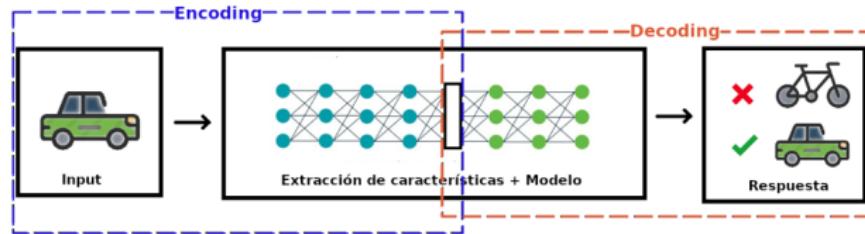
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Este concepto ha sido fundamental para el desarrollo de los métodos de aprendizaje actuales



# Conceptos básicos XL

Victor Muñiz

Generalidades

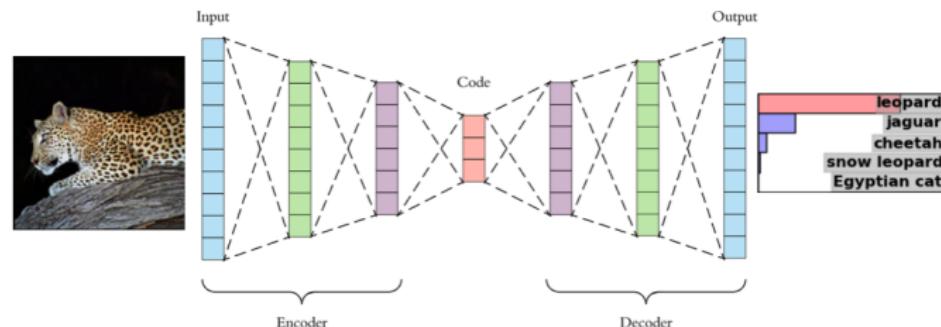
## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Conceptos básicos XLI

Victor Muñiz

## Generalidades

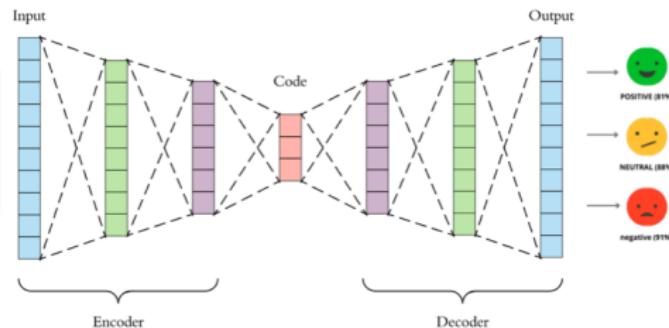
### Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Conceptos básicos XLII

Victor Muñiz

Generalidades

## Introducción

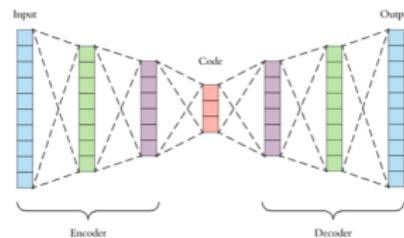
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

**“an insect robot  
preparing a delicious  
meal”**



# Conceptos básicos XLIII

Victor Muñiz

Generalidades

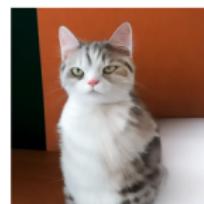
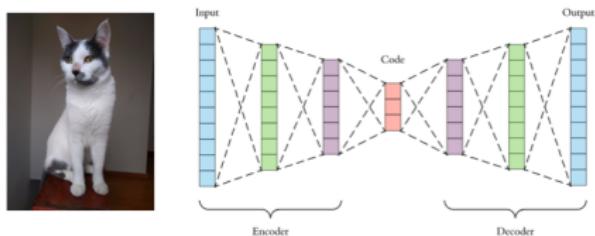
## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Conceptos básicos XLIV

Victor Muñiz

Generalidades

Introducción

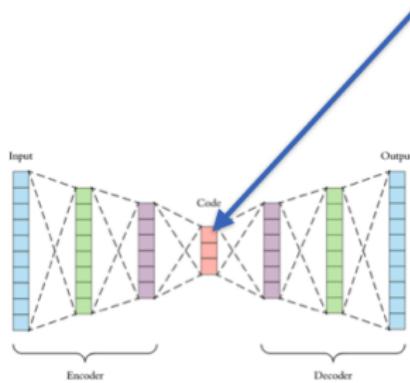
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

En ésta parte del curso, partiremos de una representación (vectorial) adecuada de los datos, y nos enfocaremos en describir modelos de predicción.



Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Conceptos básicos XLV

## Medidas de similaridad

La disimilitud o proximidad la representamos como una función de **distancia**

$$d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j),$$

y para  $n$  objetos, esta disimilitud la expresamos mediante

$$\mathbf{D}_{n \times n},$$

con  $D_{ij} = d_{ij}$ .

La función de distancia debe tener ciertas propiedades...

- $d(x, y) \geq 0$
- $d(x, y) = 0 \leftrightarrow x = y$
- $d(x, y) = d(y, x)$

# Conceptos básicos XLVI

Victor Muñiz

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

- $d(x, y) \leq d(x, z) + d(z, y)$

En general, la disimilaridad es una pseudo-distancia.

Distancias: considera el caso general

$$d_{ij}^p = \left( \sum_{k=1}^d |x_{ik} - x_{jk}|^p \right)^{1/p} \text{ (Minkowski)}$$

# Conceptos básicos XLVII

Victor Muñiz

Generalidades

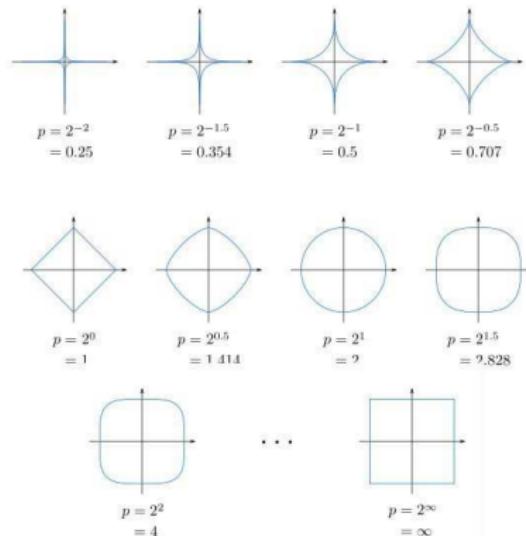
## Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



Y también:  $d_{ij} = 1 - \rho_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2$ , donde la última igualdad, se cumple solo si las observaciones están estandarizadas.

## Generalidades

## Introducción

Elementos de teoría de decisión estadística

Métricas de desempeño

Prediction-interpretability tradeoff

Bias-variance tradeoff

## Conceptos básicos XLVIII

Medidas de distancias para ML (Sahel Eskandar, [medium.com/eskandar.sahel](https://medium.com/eskandar.sahel)):

| Picture   | Method                      | Application  | Features   | Disadvantages   | Formula          |
|---|-----------------------------|--|--|---|------------------|
|  | <b>Euclidean Distance</b>   | General distance measurement, Clustering, Classification, Regression | Measures the straight line distance between two points in n-dimensional space.                       | Sensitive to outliers, Can be affected by scale differences   | $O(n)$<br>Fast   |
|  | <b>Manhattan Distance</b>   | Distance on grid networks, Routing algorithms, Image processing      | Measures the distance between two points on a grid network, where movement is limited.               | Ignores diagonal movement, not useful for high-dimensional data,  | $O(n)$<br>Fast   |
|  | <b>Cosine Similarity</b>    | Text document clustering, Text analysis, Recommendation systems      | Measures the cosine of the angle between two vectors   | Ignores magnitude of vectors, Not useful for negative values or high degree of correlation data             | $O(n)$<br>Fast   |
|  | <b>Minkowski Distance</b>   | General distance measurement   | Measures the distance between two points in n-dimensional space, where r determines the metric used. | Sensitive to outliers   | $O(n)$<br>Fast   |
|  | <b>Chebyshev Distance</b>   | Measuring maximum difference, Clustering, Anomaly detection          | Measures the maximum difference between corresponding components of two vectors                      | Only applicable for continuous data, Sensitive to outliers, may not be as useful for highly correlated data | $O(n)$<br>Fast   |
|  | <b>Hamming Distance</b>     | Measuring string similarity, Error-correcting codes, DNA sequencing  | Measures the number of positions at which the corresponding symbols are different.                   | Only for same length strings, May not be as useful for continuous data                                      | $O(n)$<br>Fast   |
|  | <b>Levenshtein Distance</b> | Measuring string similarity  | Measures the minimum number of single-character edits required to transform one string into another. | More expensive for long strings   | $O(n^2)$<br>Slow |
|  | <b>Jaccard Similarity</b>   | Set similarity measurement, Text analysis, recommendation systems    | Measures the similarity between two sets by comparing their intersection and union.                  | Ignores magnitude of sets, May not be as useful for continuous data   | $O(n)$<br>Fast   |
|  | <b>Sørensen-Dice Index</b>  | Measuring similarity of sets, Ecology, Biology, Genetics             | Measures the similarity between two sets   | May not be as useful for continuous data and Ignores magnitude of sets                                      | $O(n)$<br>Fast   |

# Conceptos básicos XLIX

Victor Muñiz

## Generalidades

### Introducción

Elementos de teoría de decisión estadística

Métricas de desempeño

Prediction-interpretability tradeoff

Bias-variance tradeoff

| Picture | Method                            | Application   | Features   | Disadvantages   | Formula               |
|---------|-----------------------------------|---|--|---|-----------------------|
|         | <b>Haversine Distance</b>         | Measuring distance on a sphere, Geographic calculations                               | Measures the great-circle distance between two points on a sphere, such as the Earth.  | Not useful for small distances where a spherical Earth is not an accurate representation            | $O(n)$<br>Slow        |
|         | <b>Mahalanobis Distance</b>       | Multivariate statistical analysis, Outlier detection, Clustering                      | Measures the distance between two points in n-dimensional space, taking into account the correlation between variables.            | Requires full covariance matrix, May not be as useful for datasets with a large number of variables | $O(n^3)$<br>Slow      |
|         | <b>Pearson Correlation</b>        | Measuring linear correlation  | Measures the linear correlation between two variables in a dataset.  | Requires linear correlation   | $O(n^2)$<br>Slow      |
|         | <b>Squared Euclidean Distance</b> | Clustering algorithms   | Measures the square of the straight line distance between two points in n-dimensional space.                                       | More sensitive to outliers  | $O(n)$<br>Fast        |
|         | <b>Jensen-Shannon Divergence</b>  | Measuring similarity of probability distributions. Clustering, Recommendation systems | symmetric measure used to compare probability distributions, considers both similarities and differences between the distributions | Only applicable for non-negative vectors  | $O(n)$<br>Slow        |
|         | <b>Chi-Square Distance</b>        | Measuring similarity of histograms  | Measures the distance between two histograms by comparing their Chi-Square divergence.   | Only applicable for non-negative vectors  | $O(n)$<br>Fast        |
|         | <b>Spearman Correlation</b>       | Measuring rank correlation  | Measures the correlation between two variables in a dataset based on their rank order.   | Only applicable for ordinal data  | $O(n \log n)$<br>Fast |
|         | <b>Canberra Distance</b>          | Measuring distance for sparse data  | Measures the distance between two vectors, taking into account their relative magnitudes.  | More sensitive to small differences   | $O(n \log n)$<br>Fast |

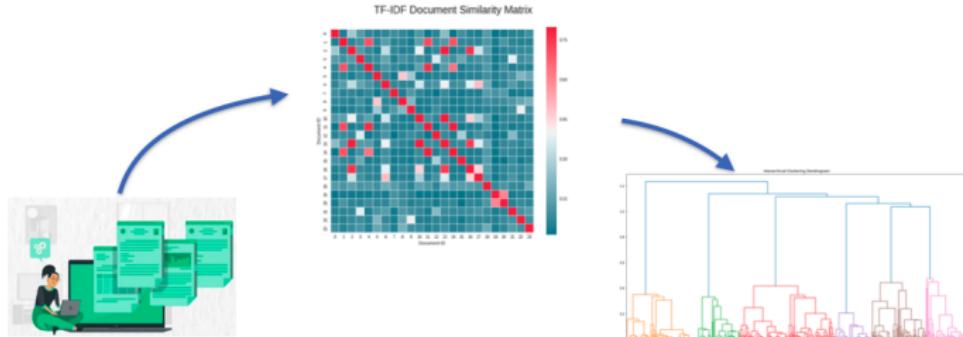
# Conceptos básicos L

Ya que tenemos definida una **representación vectorial** de los datos y una **medida de distancia** apropiada, podemos usarla para varias tareas de ML.

## Clustering

### Clustering jerárquico

Se construye a partir de una **matriz de similaridades** de los datos



# Conceptos básicos LI

Victor Muñiz

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

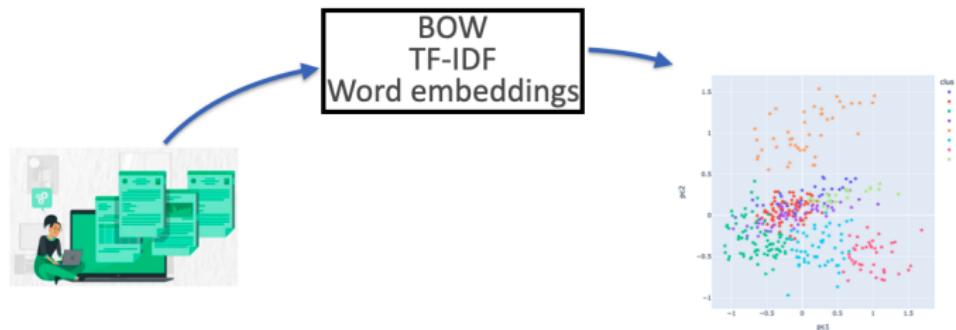
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Combinatorios: $k$ -means

Es un algoritmo combinatorio que asigna  $n$  objetos en  $k$  clústeres, de acuerdo al criterio de distancia mínima entre cada objeto y los "prototipos" de cada clúster. Generalmente, se usa el centroide como prototipo de los clústeres.



# Conceptos básicos LII

Victor Muñiz

Generalidades

Introducción

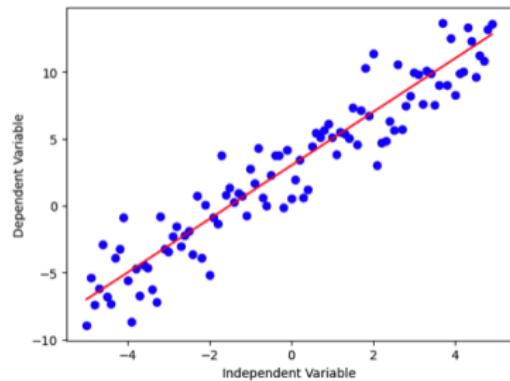
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

$$\begin{aligned}y &= f(x) \\&= \beta_0 + \beta_1 x\end{aligned}$$



# Conceptos básicos LIII

Victor Muñiz

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

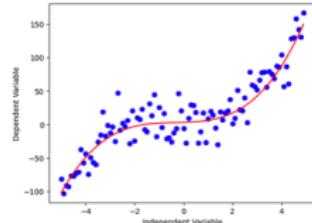
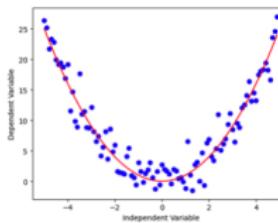
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

$$y = f(x)$$

Pero  $f$  puede ser  
más compleja...



# Conceptos básicos LIV

Victor Muñiz

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

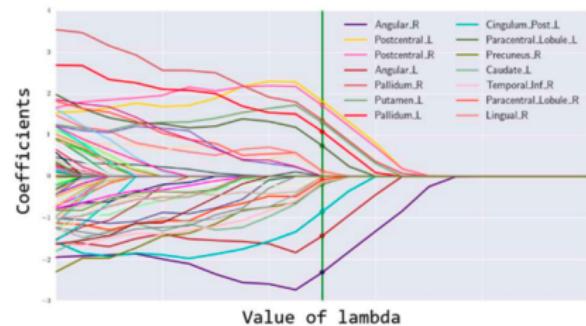
Métricas de desempeño  
Prediction-interpretability  
tradeoff

Bias-variance tradeoff

$$y = f(x) + \text{regularización}$$

Métodos sofisticados de  
regresión:

- Ridge
- Lasso
- Elastic-Net
- LARS



# Conceptos básicos LV

Victor Muñiz

Generalidades

## Introducción

Elementos de teoría de  
decisión estadística

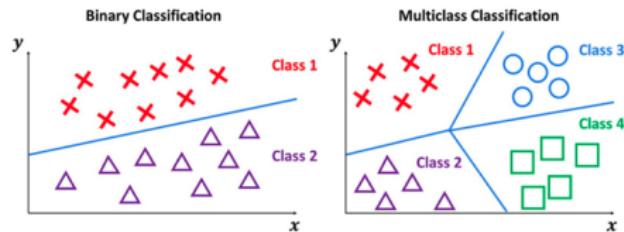
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

$$y = f(x)$$

$y$  puede tomar un  
valor discreto 1,2,...,K



En las siguientes sesiones, veremos varios de éstos métodos,  
particularmente de clasificación, y conceptos importantes  
relacionados.

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Teoría de decisión estadística

# Teoría de decisión estadística

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Consideremos un problema de clasificación binaria:

$$\{(\mathbf{x}_i, y_i)\}_{i=1}^n; \quad \mathbf{x} \in \mathbb{R}^d, y \in \{0, 1\}$$

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Teoría de decisión estadística

Supongamos que podemos obtener **de alguna forma** éstas probabilidades:

$$P(y = 0), \quad P(y = 1).$$

Si sólo contaramos con ésta información, lo más lógico (óptimo) es asignar

$$y_i = 1 \quad \text{si} \quad P(y = 1) > P(y = 0)$$

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Teoría de decisión estadística

Sin embargo, nosotros contamos con información valiosa sobre la categoría de las observaciones a través de las covariables  $\mathbf{x}$ .

Si una covariable  $\mathbf{x}$  es discriminativa, esperaríamos que su distribución esté asociada a  $y$ , entonces nos interesa conocer

$$P(\mathbf{x}|y)$$

Entonces, con toda ésta información, podemos modelar

$$P(y, \mathbf{x}) = P(y|\mathbf{x})P(\mathbf{x}) = P(\mathbf{x}|y)P(y),$$

# Teoría de decisión estadística

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

entonces (Bayes):

$$P(y = 1|\mathbf{x}) = \frac{P(\mathbf{x}|y = 1)P(y = 1)}{P(\mathbf{x})},$$

con el factor de normalización

$$P(\mathbf{x}) = P(\mathbf{x}|y = 0)P(y = 0) + P(\mathbf{x}|y = 1)P(y = 1).$$

# Teoría de decisión estadística

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

entonces (Bayes):

$$P(y = 1|\mathbf{x}) = \frac{P(\mathbf{x}|y = 1)P(y = 1)}{P(\mathbf{x})},$$

con el factor de normalización

$$P(\mathbf{x}) = P(\mathbf{x}|y = 0)P(y = 0) + P(\mathbf{x}|y = 1)P(y = 1).$$

En palabras:

$$\text{posterior} = \frac{\text{verosimilitud} \times \text{apriori}}{\text{evidencia}}$$

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Teoría de decisión estadística

Para el caso de clasificación binaria, la probabilidad de error está dada por

$$P(\text{error}|\mathbf{x}) = \begin{cases} P(y = 1|\mathbf{x}) & \text{si decidimos } y = 0 \\ P(y = 0|\mathbf{x}) & \text{si decidimos } y = 1 \end{cases},$$

y el error promedio es:

$$P(\text{error}) = \int_{-\infty}^{\infty} P(\text{error}|\mathbf{x})P(\mathbf{x})dx.$$

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Teoría de decisión estadística

Si para cada  $\mathbf{x}$ , nos aseguramos que el error que cometemos es muy pequeño, entonces la integral debe ser muy pequeña, y eso se logra al usar la **regla de decisión Bayesiana**:

$$\text{decide } y = 1 \text{ si } P(y = 1|\mathbf{x}) > P(y = 0|\mathbf{x}),$$

o de forma equivalente:

$$\text{decide } y = 1 \text{ si } P(\mathbf{x}|y = 1)P(y = 1) > P(\mathbf{x}|y = 0)P(y = 0).$$

En ambos casos

$$P(\text{error}|\mathbf{x}) = \min\{P(y = 1|\mathbf{x}), P(y = 0|\mathbf{x})\},$$

ya que la categoría correcta se asigna de acuerdo a la probabilidad máxima, y bajo este supuesto, el error es asignar la probabilidad mínima.

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Teoría de decisión estadística

En forma general, el **clasificador óptimo Bayesiano** está dado por :

$$\hat{y} = \arg \max_{y_k} P(y = y_k | \mathbf{x}),$$

donde  $\mathbf{x} \in \mathbb{R}^d$ ,  $y_k \in \{1, 2, \dots, K\}$ , y las probabilidades posteriores están dadas por la fórmula de Bayes:

$$P(y = y_k | \mathbf{x}) = \frac{P(\mathbf{x}|y = y_k)P(y = y_k)}{P(\mathbf{x})}.$$

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Teoría de decisión estadística

En forma general, el **clasificador óptimo Bayesiano** está dado por :

$$\hat{y} = \arg \max_{y_k} P(y = y_k | \mathbf{x}),$$

donde  $\mathbf{x} \in \mathbb{R}^d$ ,  $y_k \in \{1, 2, \dots, K\}$ , y las probabilidades posteriores están dadas por la fórmula de Bayes:

$$P(y = y_k | \mathbf{x}) = \frac{P(\mathbf{x}|y = y_k)P(y = y_k)}{P(\mathbf{x})}.$$

Otra forma de cuantificar la consecuencia de mi decisión es a través del **riesgo** y el costo asociado.

# Teoría de decisión estadística

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Considera ésta matriz de costos para clasificación binaria

$$\Lambda = \begin{pmatrix} 0 & \lambda_{0,1} \\ \lambda_{1,0} & 0 \end{pmatrix},$$

donde  $\lambda_{i,j}$  es el costo de clasificar un objeto de la clase  $j$  como clase  $i$ . El costo promedio asociado a clasificar un dato  $\mathbf{x}$  en la clase  $y_i$ , llamado función de Riesgo de Bayes, está dado por

$$R(y = y_i | \mathbf{x}) = \sum_{j \in \{0,1\}} \lambda_{i,j} P(y = y_j | \mathbf{x}),$$

en nuestro caso,  $R(y = 0 | \mathbf{x}) = \lambda_{0,1} P(y = 1 | \mathbf{x})$  y  $R(y = 1 | \mathbf{x}) = \lambda_{1,0} P(y = 0 | \mathbf{x})$ .

# Teoría de decisión estadística

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

La decisión óptima es elegir aquella clase que minimice el riesgo, es decir

$$\hat{y} = \arg \min_{y_k} R(y = y_k | \mathbf{x}).$$

Es fácil ver que el clasificador de Bayes corresponde con esta regla de decisión, ya que ésta implica elegir  $\hat{y} = -1$  si

$$R(y = 0 | \mathbf{x}) < R(y = 1 | \mathbf{x}),$$

o, en términos de probabilidades posteriores, si

$$\lambda_{1,0} P(y = 0 | \mathbf{x}) > \lambda_{0,1} P(y = 1 | \mathbf{x}),$$

que es equivalente al clasificador óptimo de Bayes.

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Teoría de decisión estadística

En este caso, las probabilidades posteriores estarán dadas por

$$\begin{aligned}\lambda_{0,1}P(y=1|\mathbf{x}) &= \frac{P(\mathbf{x}|y=1)P(y=1)\lambda_{0,1}}{P(\mathbf{x})}, \\ \lambda_{1,0}P(y=0|\mathbf{x}) &= \frac{P(\mathbf{x}|y=0)P(y=0)\lambda_{1,0}}{P(\mathbf{x})}.\end{aligned}$$

De aquí podemos concluir que, considerar diferentes costos de mala clasificación es equivalente a **modificar las probabilidades apriori**  $P(y=y_k)$ , y en consecuencia, cambiar las fronteras de clasificación.

## Generalidades

### Introducción

Elementos de teoría de  
decisión estadística

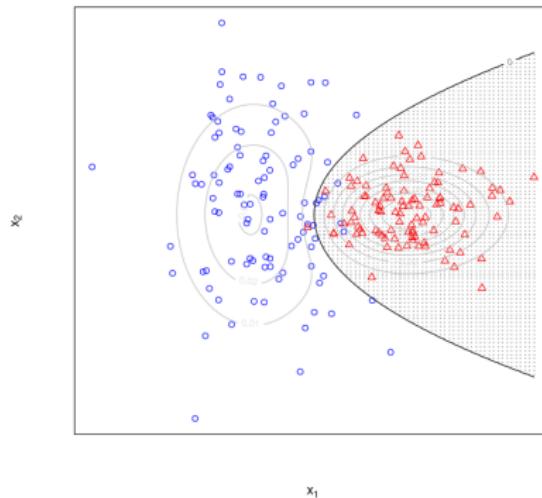
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Teoría de decisión estadística

Ejemplo: Clase 0 ( $\circ$ ). Clase 1 ( $\Delta$ ).



Clasificador Bayesiano óptimo.

$$P(y = 0) = P(y = 1) = 1/2, \quad \Lambda = \begin{pmatrix} 0 & 1 \\ 4 & 0 \end{pmatrix}$$

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

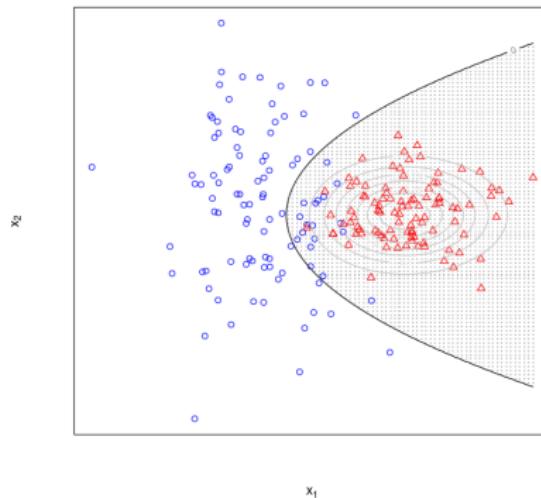
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Teoría de decisión estadística

Ejemplo: Clase 0 ( $\circ$ ). Clase 1 ( $\Delta$ ).



Clasificador Bayesiano óptimo.

$$P(y = 0) = P(y = 1) = 1/2, \quad \Lambda = \begin{pmatrix} 0 & 1 \\ 5 & 0 \end{pmatrix}$$

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

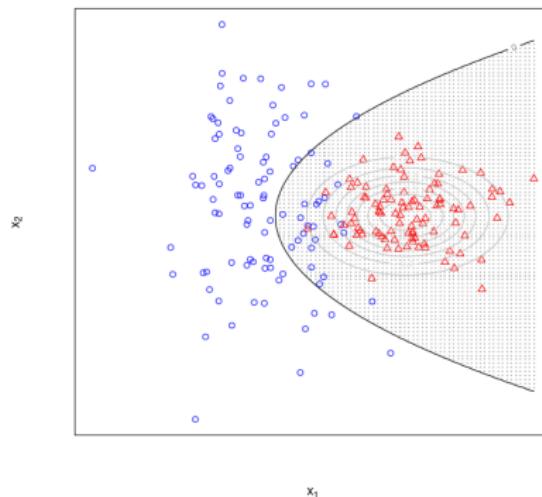
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Teoría de decisión estadística

Ejemplo: Clase 0 ( $\circ$ ). Clase 1 ( $\Delta$ ).



Clasificador Bayesiano óptimo.

$$P(y = 0) = P(y = 1) = 1/2, \quad \boldsymbol{\Lambda} = \begin{pmatrix} 0 & 1 \\ 10 & 0 \end{pmatrix}$$

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Métricas de desempeño para ML supervisado

# Métricas de desempeño I

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Regresión

MSE:

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2$$

y derivados (MAE, RMSE).

# Métricas de desempeño II

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Clasificación

Clasificación binaria:

|              |                                     | Predicted Class                                     |   |  |
|--------------|-------------------------------------|---|---|--|
|              |                                     | Positive  | Negative  |  |
| Actual Class | Positive                            | True Positive (TP)                                  | False Negative (FN)<br><b>Type II Error</b>       | Recall (sensitivity)<br>$\frac{TP}{(TP + FN)}$               |
|              | Negative                            | False Positive (FP)<br><b>Type I Error</b>          | True Negative (TN)                                | Specificity<br>$\frac{TN}{(TN + FP)}$                        |
|              | Precision<br>$\frac{TP}{(TP + FP)}$ | Negative Predictive Value<br>$\frac{TN}{(TN + FN)}$ | Accuracy<br>$\frac{TP + TN}{(TP + TN + FP + FN)}$ | $F1 = \frac{2(Precision \times Recall)}{Precision + Recall}$ |

# Métricas de desempeño III

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Accuracy (exactitud) es simplemente la proporción de datos clasificados correctamente.

$$\text{Accuracy} = \frac{TP + TN}{TOTAL}$$

donde  $TP$  es el número de predicciones positivas correctamente clasificadas,  $TN$  es el número de predicciones negativas correctamente clasificadas y  $TOTAL$  es el número total de casos u observaciones.

Es la primera métrica que se comprueba al evaluar un clasificador; sin embargo, si los datos están desbalanceados o si se está más interesado detectar una de las clases, la exactitud no captura realmente la eficacia de un clasificador.

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Métricas de desempeño IV

Precision (precisión) nos ayuda a cuantificar la fracción de positivos verdaderos entre el total de los clasificados como positivos. Refleja la habilidad de un clasificador para no arrojar falsos positivos.

$$Precision = \frac{TP}{TP + FP}$$

donde  $FP$  es el número de predicciones positivas calificadas incorrectamente (falsos positivos).

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Métricas de desempeño V

Recall (recuperación, recuerdo o sensitividad) mide la habilidad de un clasificador para detectar todas las muestras positivas.

$$Recall = \frac{TP}{TP + FN}$$

donde  $FN$  es el número de predicciones negativas calificadas incorrectamente.

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Métricas de desempeño VI

F1 score es la media armónica de la precisión y la recuperación (Recall). Esta medida es una buena forma de resumir la evaluación en un número único.

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Existe una relación entre la precisión y el recall: un modelo que predice todo como positivo tendrá un recall de 1, pero una precisión muy baja pues tendría muchos falsos positivos, mientras que un modelo que solo predijera un positivo y el resto negativos tendría un recall muy bajo, pero una precisión muy alta.

Es por ello que se recurre a la medida F1, ya que mitiga el impacto de las tasas altas y acentúa el de las tasas bajas.

# Métricas de desempeño VII

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## *Ponderación para clasificación multiclas*

- macro: esta es una media aritmética simple de todas las métricas en las clases. Esta técnica otorga el mismo peso a todas las clases, lo que la convierte en una buena opción para tareas de clasificación con clases balanceadas.
- weighted: considera el desbalanceo de las categorías calculando el promedio de métricas binarias ponderadas por el número de muestras de cada clase en el objetivo. Si 3 scores de precisión para 3 clases son: Clase 1 (0.85), clase 2 (0.80) y clase 3 (0.89), el promedio ponderado se calculará multiplicando cada puntuación por el número de datos de cada clase (soporte) y dividiendo por el número total de muestras.

# Métricas de desempeño VIII

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

- micro: es lo mismo que Accuracy. El micro-promedio se encuentra dividiendo la suma de la diagonal de la matriz de confusión por la suma de todas las celdas, es decir, la exactitud (accuracy). Como ésta métrica no es recomendable si se tienen clases muy desbalanceadas, esta forma de promediar rara vez se usa.

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## El compromiso entre precisión (exactitud) e interpretabilidad

# Prediction-interpretability I

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Una de las características de los modelos de ML para predicción es su **flexibilidad**, que está relacionado con el concepto de complejidad.

Entendemos como flexibilidad, la capacidad de un modelo para generar distintas “formas” de  $f$ .  
Observa los siguientes ejemplos:

# Prediction-interpretability II

Victor Muñiz

Generalidades

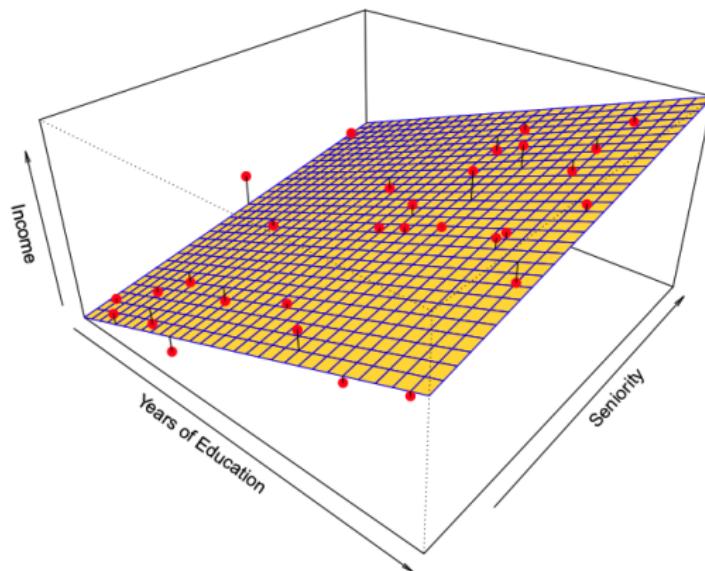
Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Prediction-interpretability III

Victor Muñiz

Generalidades

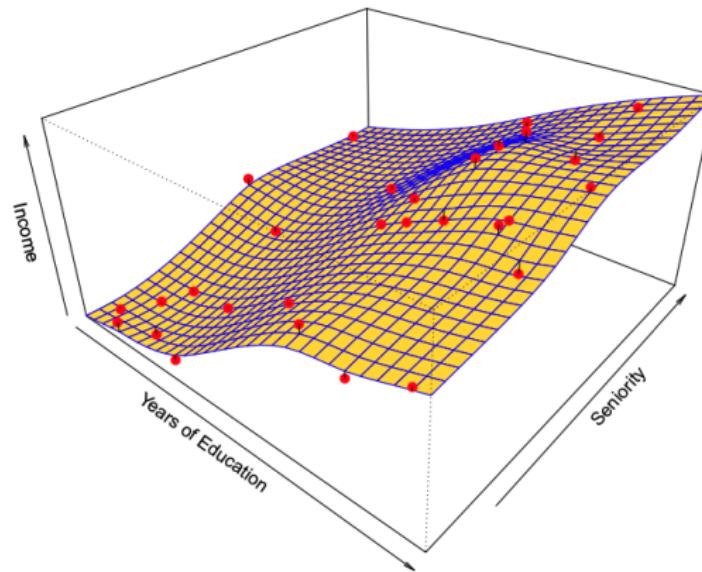
Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Prediction-interpretability IV

Victor Muñiz

Generalidades

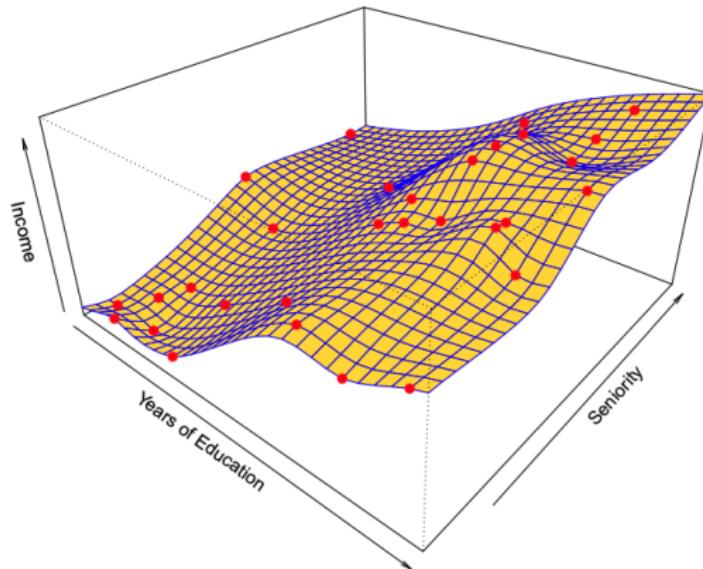
Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



¿Porqué escoger un modelo más restrictivo o poco flexible?

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Prediction-interpretability V

## Interpretabilidad

Consideramos que un modelo es interpretable si podemos inferir las relaciones entre las variables independientes  $x$  y la dependiente  $y$ .



# Prediction-interpretability VI

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Es muy complicado tener ambas cosas: un modelo preciso e interpretable.

Actualmente, es un tema de investigación muy importante en ML.

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## El compromiso entre sesgo y varianza

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

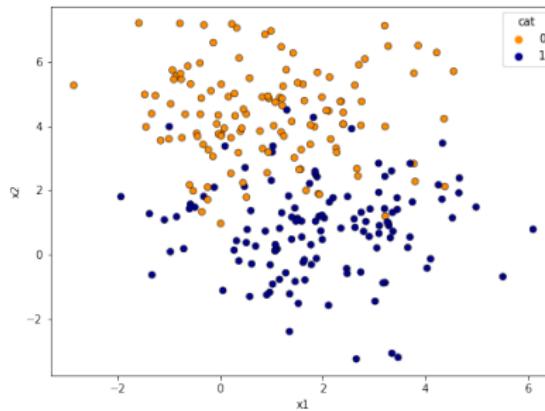
Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Bias-variance tradeoff I

Considera éste esquema clásico de clasificación.



Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

## Bias-variance tradeoff II

Donde tenemos un conjunto de datos de entrenamiento

$$\{(\mathbf{x}_i, y_i)\}_{i=1}^n; \quad \mathbf{x} \in \mathbb{R}^d, y \in \{0, 1\}$$

Nuestro objetivo es obtener una función

$$f : \mathbb{R}^d \mapsto \{0, 1\}.$$

Bajo el esquema de aprendizaje máquina (ML), esperamos que ésta función se “aprenda” a partir de un conjunto de datos de entrenamiento, y generalmente depende de ciertos parámetros:

$$y = f(\mathbf{x}; \boldsymbol{\theta}).$$

¿Cómo nos gustaría que fuera  $f$ ?

# Bias-variance tradeoff III

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Veamos dos modelos para resolverlo.

Modelo lineal

$$f(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}' \mathbf{x}$$

$$\hat{y} = \begin{cases} 0 & \text{si } f(\mathbf{x}) \leq 0.5 \\ 1 & \text{si } f(\mathbf{x}) > 0.5 \end{cases}$$

# Bias-variance tradeoff IV

Victor Muñiz

Generalidades

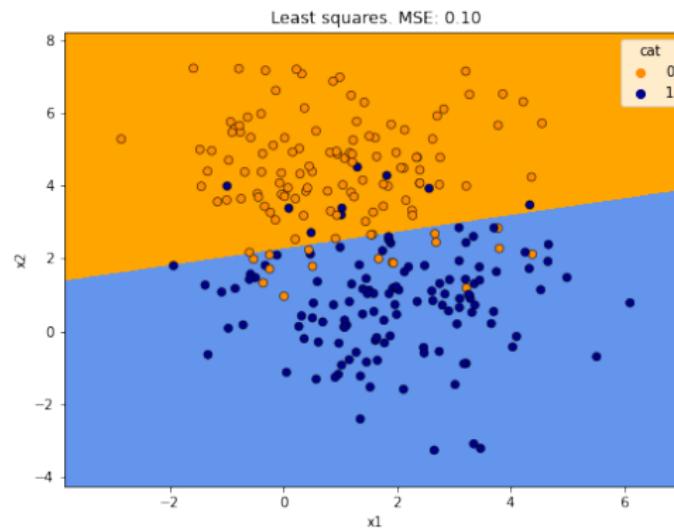
Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

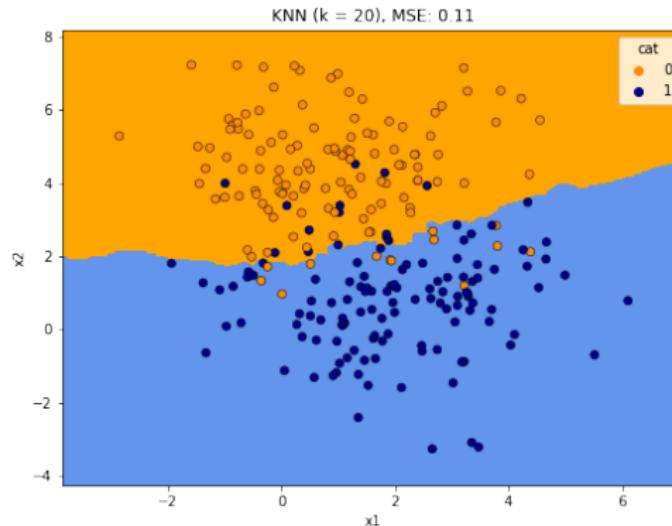
Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Bias-variance tradeoff V

$k$ -vecinos cercanos:

$$\hat{y} = f(\mathbf{x}) = \frac{1}{k} \sum_{\mathbf{x}_i \in \mathcal{N}_k(\mathbf{x})} y_i$$



# Bias-variance tradeoff VI

$k$ -vecinos cercanos:

$$\hat{y} = f(\mathbf{x}) = \frac{1}{k} \sum_{\mathbf{x}_i \in \mathcal{N}_k(\mathbf{x})} y_i$$

Generalidades

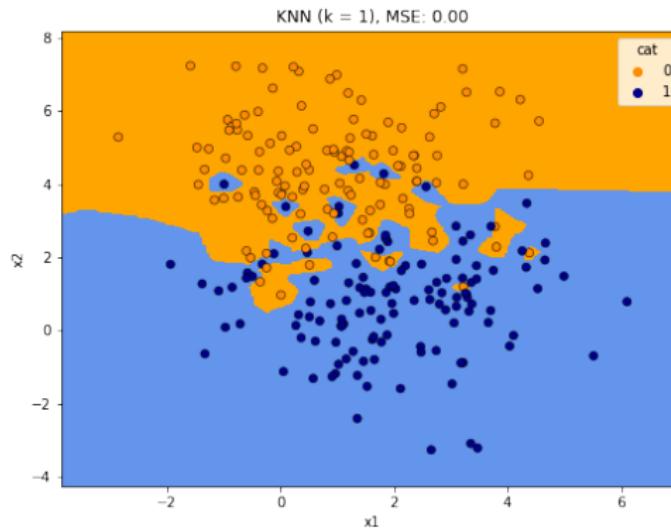
Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



# Bias-variance tradeoff VII

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

¿Cuál prefieres y porqué?

# Bias-variance tradeoff VIII

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

Observa que el error esperado de **cualquier** modelo de predicción en una observación dada  $\mathbf{x}_0$  puede descomponerse mediante:

$$\begin{aligned}\text{error}(\mathbf{x}_0) &= E((y - \hat{f}(\mathbf{x}_0))^2) \\ &= \sigma_{\epsilon}^2 + \text{Bias}^2(\hat{f}(\mathbf{x}_0)) + \text{Var}(\hat{f}(\mathbf{x}_0)) \\ &= \text{Error} + \text{Bias}^2 + \text{Variance}.\end{aligned}$$

# Bias-variance tradeoff IX

Victor Muñiz

Generalidades

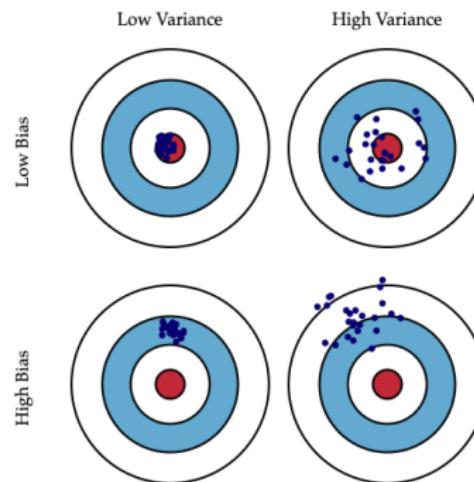
Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



¿Cómo nos sirve esto para determinar el modelo “adecuado” ?  
Bias-Variance tradeoff

# Bias-variance tradeoff X

Victor Muñiz

Generalidades

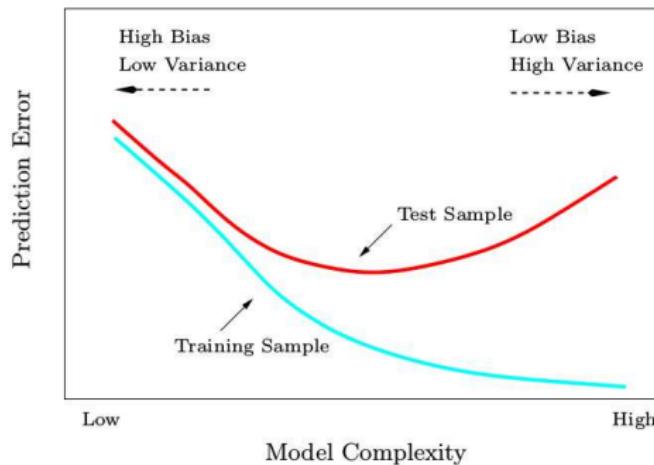
Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff



Hastie, et al. 2nd. Ed.

Generalidades

Introducción

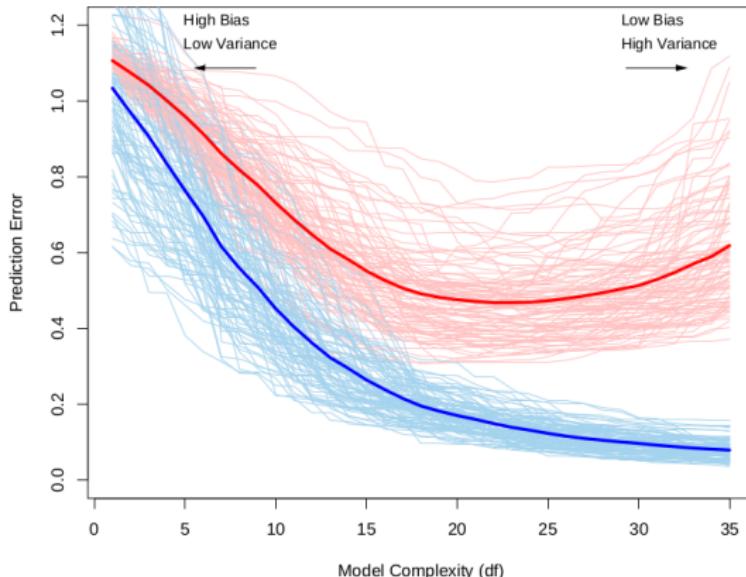
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Bias-variance tradeoff XI



Hastie, et al. 2nd. Ed.

Bias-Variance tradeoff esquemáticamente

Generalidades

Introducción

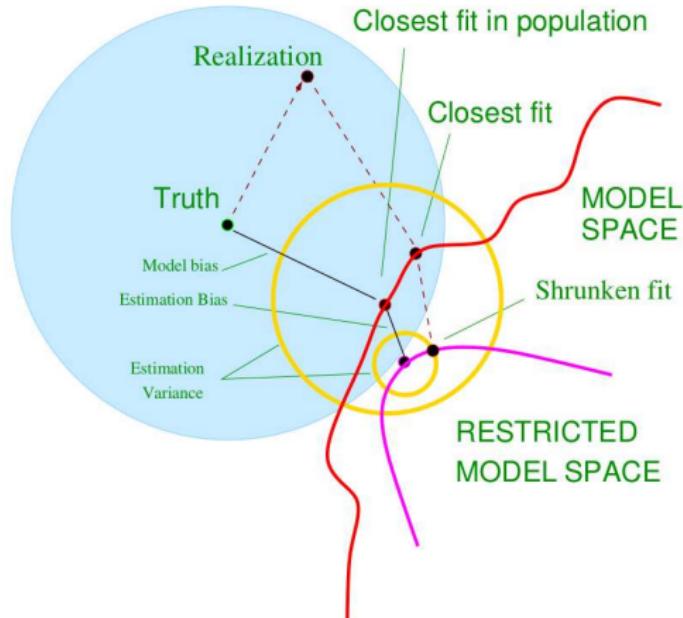
Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

# Bias-variance tradeoff XII



Hastie, et al. 2nd. Ed.

Nuestros objetivos.

# Bias-variance tradeoff XIII

Victor Muñiz

Generalidades

Introducción

Elementos de teoría de  
decisión estadística

Métricas de desempeño

Prediction-interpretability  
tradeoff

Bias-variance tradeoff

- Selección del modelo: estimación del desempeño de diferentes modelos para escoger el mejor.
- Evaluación del modelo: una vez escogido un modelo (final), estimar su error de predicción (error de generalización) en un nuevo conjunto de datos.