Metodologías de análisis: machine learning y visión general

Introducción a big data

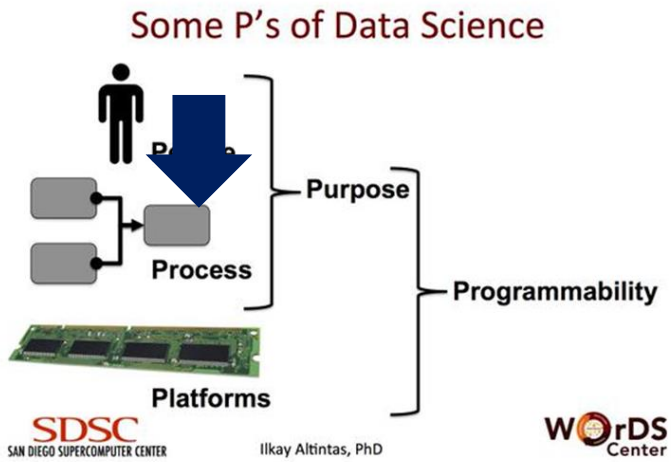
Prof. César Moreno Pascual

http://es.linkedin.com/in/cesarmorenopascual/

Universidad Francisco de Vitoria
UFV Madrid

Índice

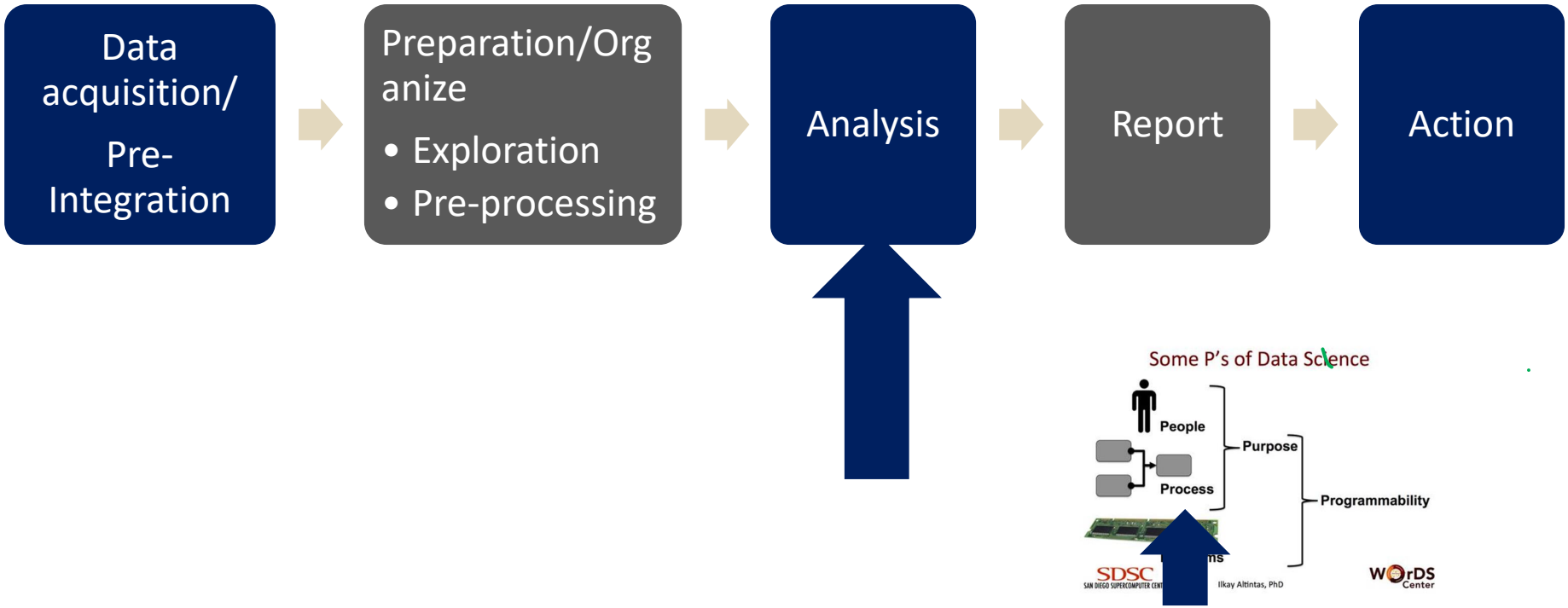
- Introducción
- supervisado
  - Predicción
  - Clasificación
- sin supervisión
- Los datos no estructurados y estructurados
- Aprendizaje profundo
- otras perspectivas
  - Los sistemas de recomendación
  - análisis de redes



5 Sal de Big Data

Procesar: Recuerde. Coste, plan, paquetes de trabajo, los entregables, ...

Es un proyecto de I + D



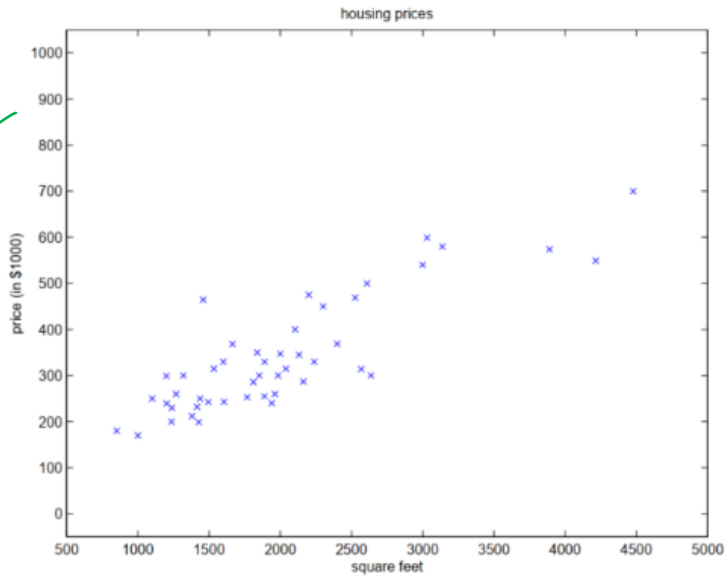
Introducción nomenclatura

31/1/25

precio relacionado con los m^2.

Living area (feet^2)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮

precio



m^2

relación real

$Y = f(x) + \varepsilon \Rightarrow \hat{Y} = \hat{f}(x)$

precio — Error cero — suponiendolo.

La variable de entrada (X) (también conocida n como **predictor, variable independiente, característica o simplemente variable**) es. Cada una de las muestras de la entrada se denota como $\mathbf{x^{(i)}}$

Entonces la variable de salida (Y) (precio) (**también conocida como respuesta,** variable **dependiente o variable** objetivo) es el precio. Cada una de las muestras del output se denota como $\mathbf{y^{(i)}}$**.**

Introducción nomenclatura

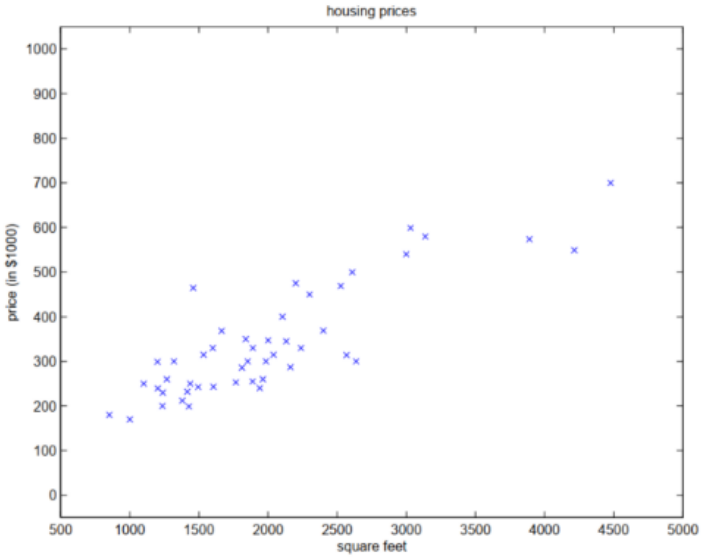
$$E(y-\hat{y})^2 = \left(f(x)+\varepsilon-\hat{f}(x)\right)^2 = \left(f(x)-\hat{f}(x)\right)^2 + \left(\varepsilon^2\right)$$

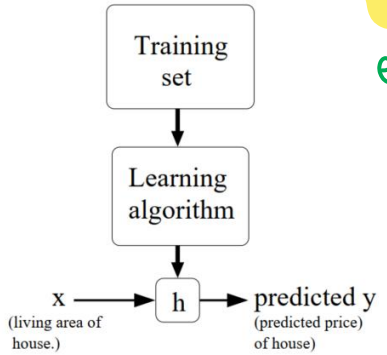
error reducible = 0

varianza del error.

error aleatorio

Living area (feet^2)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮





el problema de aprendizaje supervisado un poco más formalmente, nuestro objetivo es, dado un conjunto de entrenamiento, aprender una función h: X → Y para que **h (x)** sea un predictor "bueno" para el valor correspondiente de y. Por razones históricas, esta función h se llama hipótesis. Visto pictóricamente, el proceso es así (Ng, 2012):

Introducción nomenclatura

31/1/23

Imaginamos que hacemos → $Precio = \beta_0 + \beta_1 m^2$

¿Cuál es la inferencia?

¿Qué es la inferencia?

→ Es el porqué de las predicciones

¿Cuánto varía el precio cuando sube el precio?

Métodos → inferencia

Redes neuronales → buena predicción

¿Qué es el aprendizaje supervisado?

Entrenar el algoritmo para encontrar f. o $\hat{f}$

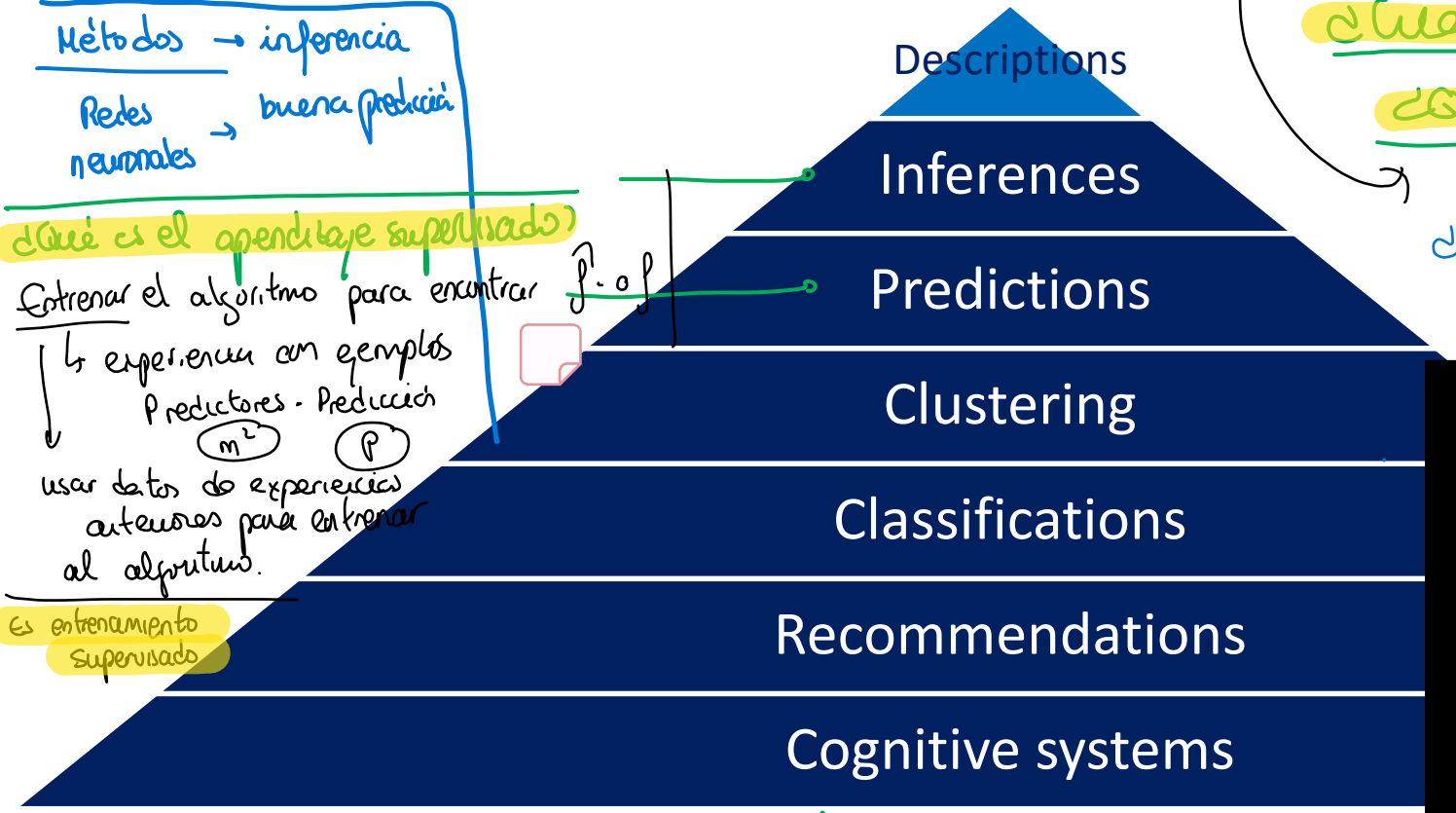
↳ experiencia con ejemplos

Predictores - Predicción

(m^2) (P)

usar datos de experiencias anteriores para entrenar al algoritmo.

Es entrenamiento supervisado



Diferencia entre predicción y clasificación

Predicción → variable continua

Clasificación → variable discontinua

- **Regresión (predicción)**
  - **<u>Predicción</u>**: Ya que el término **0 promedios podemos predecir $\hat{f}$ usando como una caja negra:**

$$\hat{Y} = \hat{f}(X)$$

  - **La predicción de Y depende de dos cantidades: $\hat{Y}$**
    - **error reducible**
      - Depende de la precisión de $\hat{f}$
    - **irreductible de error**
      - Recuerde que Y depende también de Ɛ que no puede predecirse utilizando X
        - El error no medidos contiene variables

$$\begin{aligned} E(Y - \hat{Y})^2 &= E[f(X) + \epsilon - \hat{f}(X)]^2 \\ &= \underbrace{[f(X) - \hat{f}(X)]^2}_{\text{Reducible}} + \underbrace{\text{Var}(\epsilon)}_{\text{Irreducible}}, \end{aligned}$$

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres

Aprendizaje supervisado: inferencia

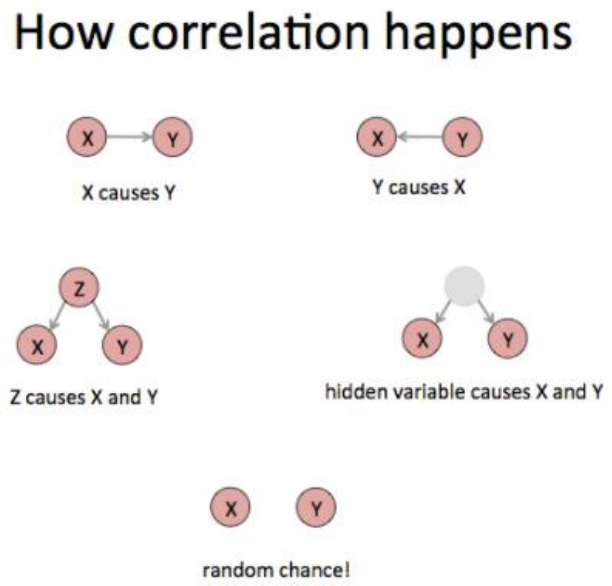
- **Regresión (Inferencia)**
  - Inferencia: Estamos interesados en la comprensión de la forma en que Y es afectado como X: X_1.....X_{pag} cambio

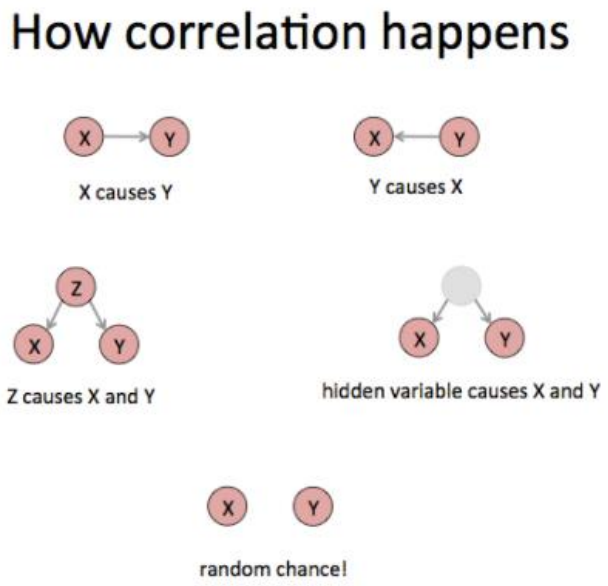
$$\hat{Y} = \hat{f}(X)$$

  - **Ahora $\hat{f}$ no es una caja negra porque necesitamos conocer su forma exacta:**
    - Los predictores que están asociados con la respuesta
    - Cuál es la relación entre la respuesta y cada predictor

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
elaboración propia

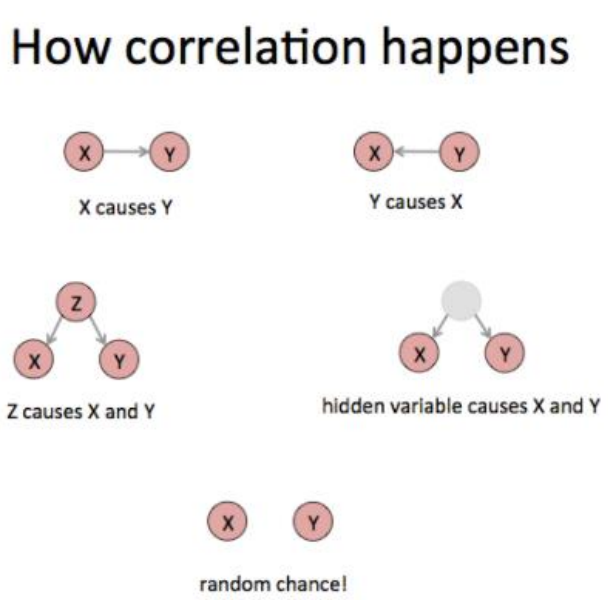
- Si bien la causalidad y la correlación pueden existir al mismo tiempo, la correlación no implica causalidad. **La causalidad** se aplica explícitamente a los casos en que la acción X causa el resultado Y
- La correlación y la causalidad a menudo se confunden porque a la mente humana le gusta encontrar patrones incluso cuando no existen

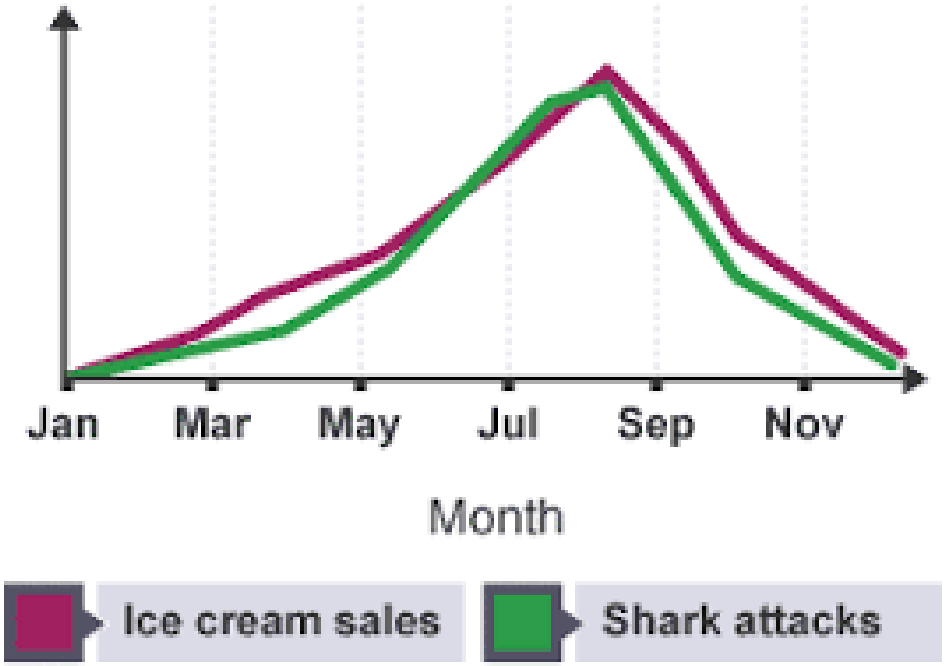




- X causa Y
- Lo contrario es cierto: Y causa X.
- Los dos están correlacionados, pero hay más: X e Y están correlacionados, pero son causados por Z.
- Hay otra variable involucrada: X causa Y, siempre y cuando Z suceda.
- Hay una reacción en cadena: X causa Z, lo que lleva a Z a causar Y (pero solo viste que X causa Y de tus propios ojos).
- Correlación casual

Aprendizaje supervisado: inferencia





Aprendizaje supervisado: enfoque sparamétricos

- **Regresión (-Inferencia Predicción): Estimación de parámetros**
  - **¿Cómo calculamos $\hat{f}$: métodos paramétricos**
    - **Paso 1: se asume una forma de la función**

$$Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p.$$

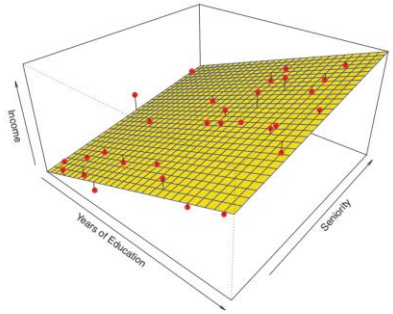


FIGURE 2.4. *A linear model fit by least squares to the* Income *data from Figure 2.3. The observations are shown in red, and the yellow plane indicates the least squares fit to the data.*

    - ***Paso 2: utilizamos los datos de entrenamiento para adaptarse o entrenar el modelo***
      - ***Calculamos los parámetros***
        - mínimos cuadrados ordinarios: descenso de gradiente descendente
  - **El modelo que elegimos por lo general no coincidirá con la verdadera desconocida $\hat{f}$**
    - ***Podemos elegir modelos flexibles que pueden caber muchas formas funcionales diferentes, pero:***
      - Los supone más complejos para el cálculo de varios parámetros
      - Puede conducir a sobreajuste

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
elaboración propia

Aprendizaje supervisado: métodos no paramétricos

- **Regresión (-Inference predicción)**
  - **¿Cómo calculamos $\hat{f}$ métodos no paramétricos**

    No se hacen suposiciones explícitas acerca de la forma de función de $\hat{f}$

    - Buscamos para estimar f para que llegue lo más cerca de los puntos de datos sin ser demasiado áspera o ondulada

  - Dado que no reducimos el cálculo para un pequeño número de parámetros, **necesitamos un gran número de observaciones**

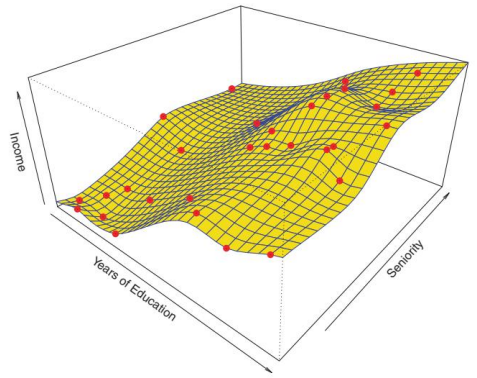


FIGURE 2.6. *A rough thin-plate spline fit to the* Income *data from Figure 2.3. This fit makes zero errors on the training data.*

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
elaboración propia

Aprendizaje supervisado

- **Regresión (-Inference predicción)**
  - **precisión de la predicción y la interpretación Modelo**
    - Algunos **métodos menos flexibles** como regresión lineal puede producir una gama de formas relativamente pequeña para estimar $\hat{f}$
    - Otros, **mas flexibles**, Como splines pueden generar una **más amplia gama de posibles formas**

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
elaboración propia

Aprendizaje supervisado: parmétricos ajuste de la función

Living area (feet^2)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮

$$f_{w,b}(x) = b + b\, w_1$$

Aquí, las $\mathbf{w_i}$ son los parámetros (también llamados pesos) que parametrizan el espacio de funciones lineales que mapean de X a Y.

Cómo de cerca están las $f_{w,b}(x^i)$ **'s** de las $\mathbf{y^{(i)}}$ correspondientes

Queremos elegir w para minimizar **J(w).**

$$J(w, b) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

Aprendizaje supervisado: parmétricos ajuste de la función

Living area (feet^2)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮

$$f_{w,b}(x) = b + b\,w_1$$

Queremos elegir w para minimizar **J(w).**

$$J(w, b) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$w_j := w_j - \alpha \frac{\delta}{\delta w_j} J(w, b)$$

regla de aprendizaje Widrow-Hoff

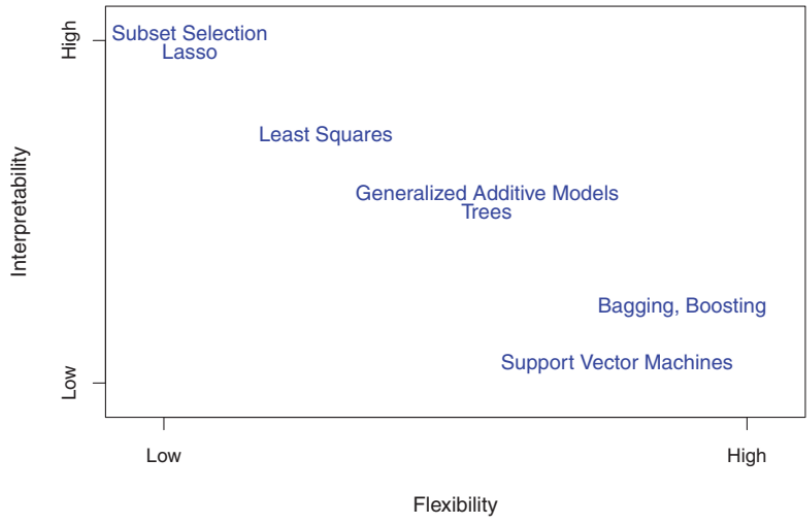
repeat until convergence {

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x^{(i)}$$

}

- **Regresión (-Inference predicción)**
  - **P-flexibilidad e la interpretabilidad del Modelo**
    - ***¿Por qué optar por utilizar un método más restrictivo en lugar de un enfoque muy flexible?***
      - ***Por inferencia:*** más flexible son menos interpretable



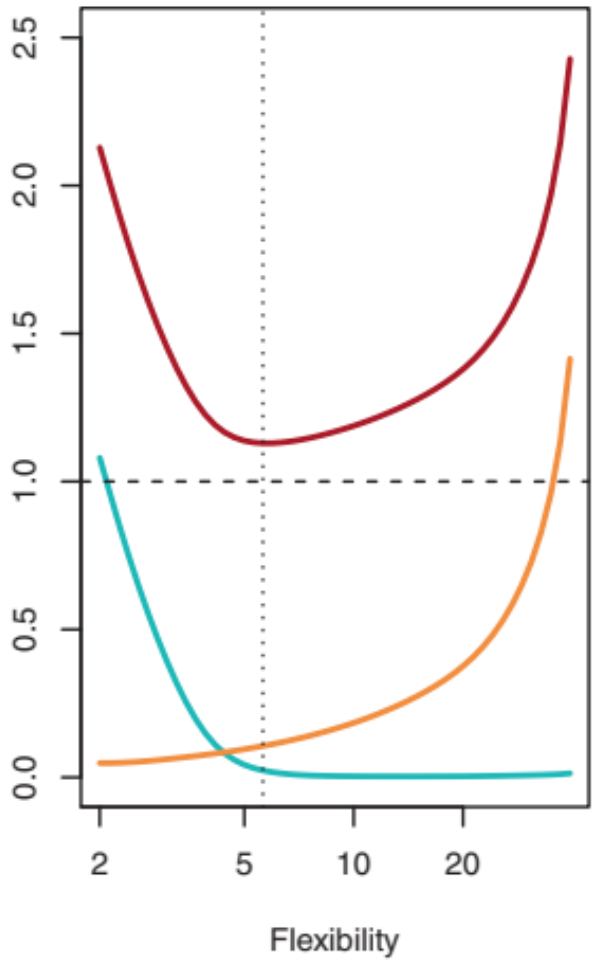
      - ***Para la predicción:*** a menudo se obtienen predicciones más precisas utilizando un métodos menos flexibles

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
elaboración propia

- **predicción -Inference**
  - **precisión de la predicción y la interpretación Modelo**
    - ***Disyuntiva de varianza y sesgo (variance-bias)***
      - **Varianza**: se refiere a **cantidad por la cual cambiaría** Si calculamos que el uso **diferente conjunto de datos (la forma de no cambia)**
        - Dado que los datos de entrenamiento se u$\hat{f}$izan para adaptarse a la, **diferentes conjuntos $\hat{f}$e datos dará lugar a una diferente**
        - Lo ideal sería que la estimación no varían demasiado, pero, si un método tiene **alta varianza** a continuación, los pequeños cambios pueden resultar en **grandes cambios**
      - ***Sesgo:*** se refiere al error que se introduce por si $\hat{f}$ **se aproxima la función a la real o no (se modifica la función)**
    - **En los métodos más flexibles, la varianza se incrementará y el sesgo disminuirá**

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
elaboración propia

- **predicción -Inferencia**
  - ***Disyuntiva de varianzas y sesgo***
  - **métodos más flexibles la varianza aumentarán y el sesgo disminuirán**
    - **Naranja: varianza (debido a que el cambio en los datos de entrenamiento)**
    - **Azul: sesgo (por el tipo de modelo)**
    - **Red: error Least Square (medida de la exactitud del método)**



Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres

- **Regresión Lineal models_Prediction e inferencia**
  - **Regresión lineal simple**

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

$h_\Theta = \Theta_0 + \Theta_1 X,$

  - **coeficientes Estimación: mínimos cuadrados, de descenso de gradiente**

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$j = 0 : \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)$$

$$j = 1 : \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \cdot x^{(i)}$$

  - **La exactitud del método: alternativas**
    - **Prueba de hipotesis**
      - **H_0:** no existe una relación entre el predictor y la respuesta
      - **H_1:** hay relación entre el predictor y la respuesta
        - Calculamos el **t-estadística**. La distribución t es la probabilidad de observar el valor t o más grande suponiendo que el parámetro del modelo de cero. Esta probabilidad es el valor p, si p-valor es muy pequeño, entonces el modelo es ok

$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)}$$

    - **R^2:** proporciona la proporción de la varianza explicada tomando un valor entre 0 y 1

$$R^2 = \frac{\mathrm{TSS} - \mathrm{RSS}}{\mathrm{TSS}} = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}}$$

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres

Aprendizaje supervisado

- **Regresión: Extensión de Linear models_Predicción e inferencia**
  - **regresión lineal múltiple: predictores múltiples**

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p.$$

  - **Predictores cualitativos**

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is female} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is male.} \end{cases}$$

  - **Extensiones del modelo lineal**
    - **interacciones**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon.$$

    - **Las relaciones no lineales**

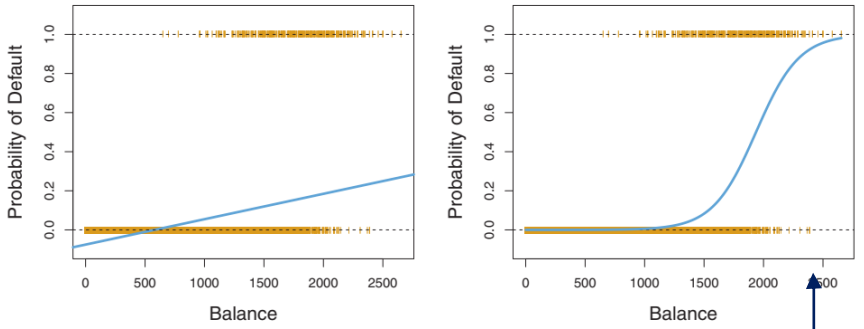
$$\texttt{mpg} = \beta_0 + \beta_1 \times \texttt{horsepower} + \beta_2 \times \texttt{horsepower}^2 + \epsilon$$

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres

- **Los modelos lineales: Clasificación**

- **Regresión logística: para clasificar una respuesta 0 o 1 de regresión lineal no es adecuada, por lo que modelar la probabilidad de estar en un grupo o el otro en lugar**

$$p(X) = \beta_0 + \beta_1 X.$$



- **Sin embargo, este enfoque no es lo suficientemente sensible porque cae todo en sí o no. así que tener una respuesta continua que finalmente calculamos**

$$\hat{y} = \sigma(w^T x + b), \quad \sigma(z) = \frac{1}{1+e^{-z}}$$

- **La estimación de los coeficientes: no hay mínimos cuadrados, pero otra función de pérdida y de descenso de gradiente diferente. En realidad el proceso es similar a lo estudiado en predicción e inferencia**

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})$$

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
Andrew Neg_deeplearning.ie
elaboración propia

Aprendizaje supervisado

- **Clasificación: clasificador Naïve Bayes**

- **Dejar que se aplica este método para la clasificación de texto. La probabilidad de un documento estar e**

$$\hat{c} = \max_{c \in C} \Pr(c) \prod_{i=1}^{n} \Pr(f_i \mid c)$$

  - **Ordenador personal$_{yo}$) Es la probabilidad de un documento conjunto de entrenamiento es en la clase c$_{yo}$. Para calcular P (c$_{yo}$):**

$$\Pr(c_i) = \frac{\text{number of docs of class } c}{\text{total number of docs in training dataset}} = \frac{N_c}{N_{docs}}$$

  - **PAG(w$_{yo}$| c$_{yo}$) Es la fracción de veces palabra w$_{yo}$aparece en todos los documentos de clase ci. En primer lugar, se crea una V vocabulario de palabras unto de entrenamiento**

$$\Pr(w_i|c) = \frac{\text{number of times } w_i \text{ appears in docs of class c}}{\text{total number of words in class c in training dataset}} = \frac{count(w_i, D_c)}{\sum_{w' \in V} count(w', D_c)} = \frac{count(w_i, D_c)}{\sum_{d \in D_c} len(d)} \quad \text{more intuitive sum}$$

$$\hat{P}(t|c) = \frac{T_{ct} + 1}{\sum_{t' \in V}(T_{ct'} + 1)} = \frac{T_{ct} + 1}{(\sum_{t' \in V} T_{ct'}) + B'}$$

Introducción a la recuperación de la información
elaboración propia

Aprendizaje supervisado

- otra clasificación methods_Classification
    - Multi-Regresión logística: respuesta binaria con múltiples predictores
    - La regresión logística con más de 2-clases
    - Naïve Bayes clasificador
    - El análisis discriminante lineal
    - K-NN (nearest neighbours) (vecinos más cercanos)
    - Árboles
    - Redes neuronales

Aprendizaje supervisado

- **métodos no lineales: predicción y clasificación**

  - **regresión polinómica**

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \ldots + \beta_d x_i^d + \epsilon_i,$$

  - **splines de regresión: En lugar de un polinomio de alto grado se ajustan varios de de bajo grado y se "suavizan" las conexiones**

$$y_i = \begin{cases} \beta_{01} + \beta_{11} x_i + \beta_{21} x_i^2 + \beta_{31} x_i^3 + \epsilon_i & \text{if } x_i < c; \\ \beta_{02} + \beta_{12} x_i + \beta_{22} x_i^2 + \beta_{32} x_i^3 + \epsilon_i & \text{if } x_i \geq c. \end{cases}$$

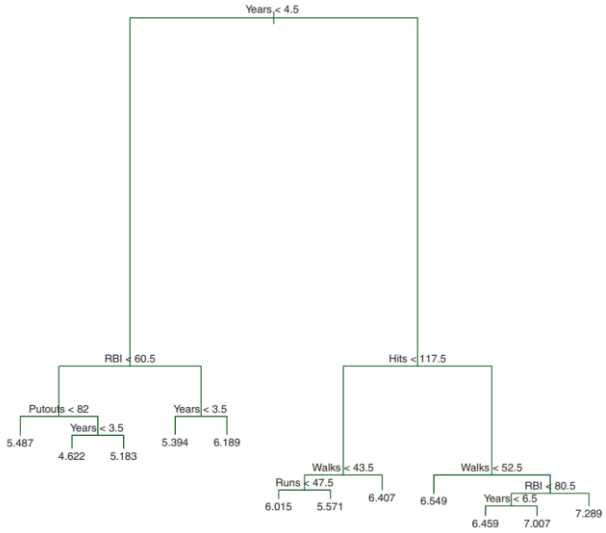
  - **modelos aditivos generalizados (GAM: General Aditive models): marco general para la ampliación de un modelo lineal. Ahora los predictores son funciones**

$$\begin{aligned} y_i &= \beta_0 + \sum_{j=1}^{\cdot} f_j(x_{ij}) + \epsilon_i \\ &= \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip}) + \epsilon_i. \end{aligned}$$

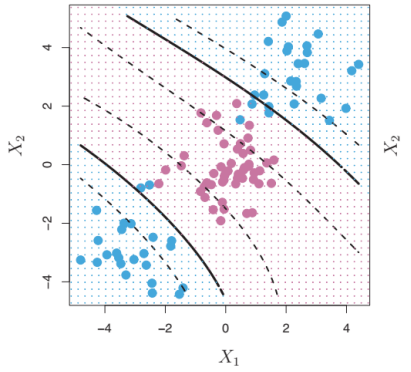
Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
elaboración propia

Aprendizaje supervisado

- **Métodos basados en árboles: predicción y clasificación**
  - **La construcción de un árbol (aprox. Método)**
    - Dividimos el predictor en regiones superpuestas en forma de nudo
    - Por cada observación que cae en la misma región hacemos la misma predicción
    - Aplicamos una función de coste
    - Repetimos hasta que la división es óptima



- **Máquinas de Vectores Soporte (Support Vector Machines): Clasificación**
  - Es una generalización del clasificador
  - Clasifica utilizando hiperplanos



Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
elaboración propia

Aprendizaje sin supervisión

- **Componentes principales**
  - El enfoque implica **la construcción de los componentes principales** y luego el uso de estos **componentes como predictores en un modelo de regresión lineal** utilizando el ajuste por mínimos cuadrados
  - A menudo, un pequeño número de componentes es suficiente para explicar la mayor parte de la variabilidad
  - Asumimos **las direcciones en que los predictores X muestran más variación** son las que **están asociados con Y**

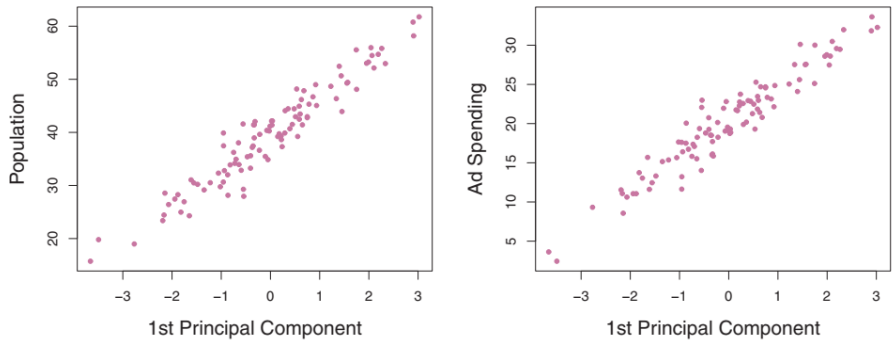


FIGURE 6.16. *Plots of the first principal component scores z_{i1} versus* pop *and* ad*. The relationships are strong.*

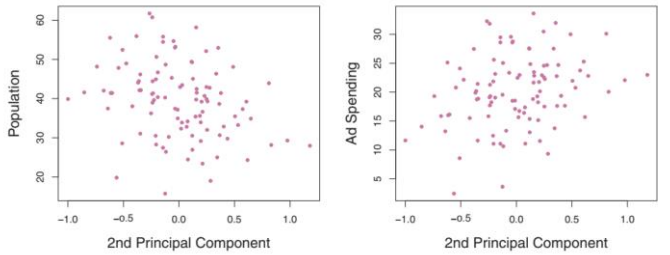


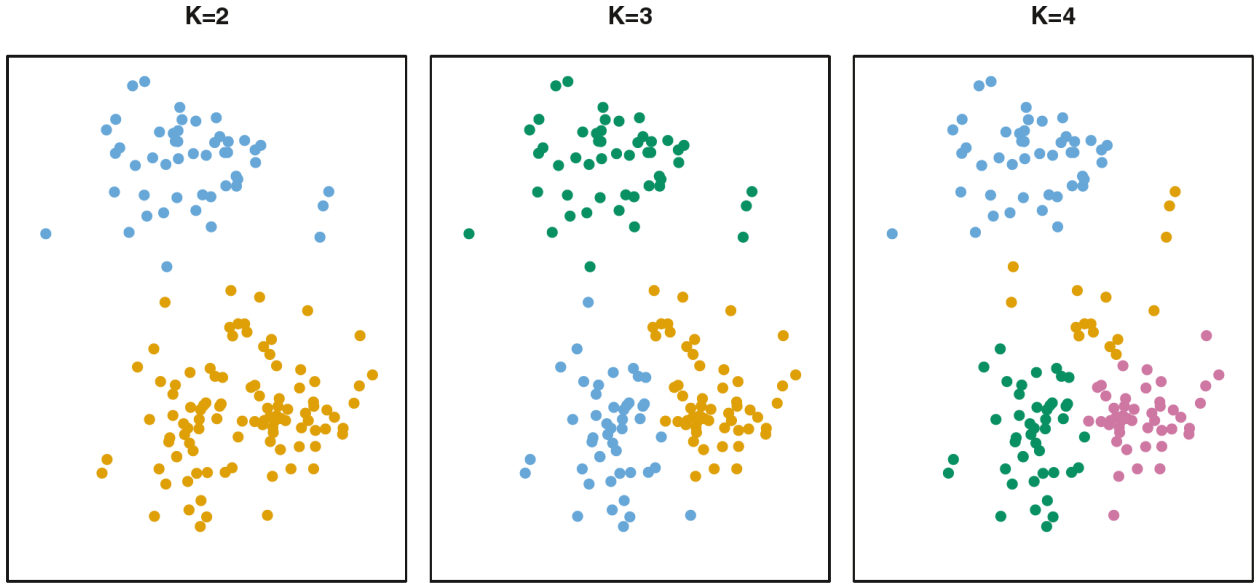
FIGURE 6.17. *Plots of the second principal component scores z_{i2} versus* pop *and* ad*. The relationships are weak.*

En los datos de publicidad, el primer componente principal explica la mayor parte de la varianza en tanto el pop y el anuncio, por lo que un director de regresión de componentes que utiliza esta sola variable para predecir alguna respuesta de interés, tales como ventas, probablemente llevará a cabo bastante bien.

Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
elaboración propia

Aprendizaje sin supervisión

- **Clusterización**
  - La clusterización se refiere a un conjunto muy amplio de técnicas para encontrar subgrupos o clústere o comunidades, en un conjunto de datos dado del que no tenemos más que la salida. Esto es, sin poder relacionar una entrada y una salida del modelo.
  - Cuando agrupamos las observaciones de un conjunto de datos, tratamos de dividirlos en grupos distintos para que las observaciones dentro de cada grupo sea similares entre sí, mientras que las observaciones en diferentes grupos sean diferentes entre sí
  - **Debemos definir qué quiere decir que dos o más observaciones sean similares o difere**

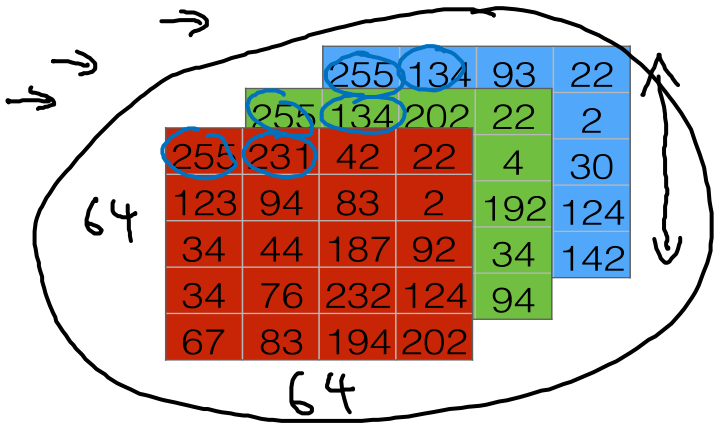


Una introducción al aprendizaje estadístico _ James G., Witten, D.,Hastie, T, Tibshirani R._Springer Nuevo yor Hiedelberg Dordrecht Londres
elaboración propia

- **Datos no estructurados: modelo de espacio vectorial**

1 (cat) vs 0 (no cat)

y



$$x = \begin{bmatrix} 255 \\ 231 \\ \vdots \\ \vdots \\ 255 \\ 134 \\ \vdots \end{bmatrix}$$

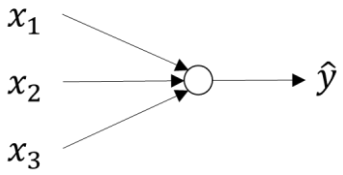
$64 \times 64 \times 3 = 12288$

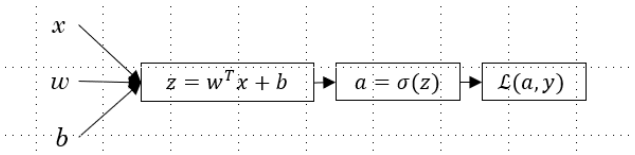
$n = n_x = 12288$

$x \longrightarrow y$

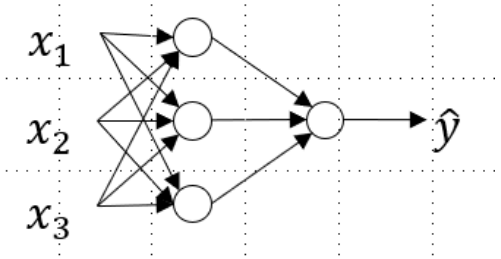
Andrew Ng_deeplearning.ie
elaboración propia

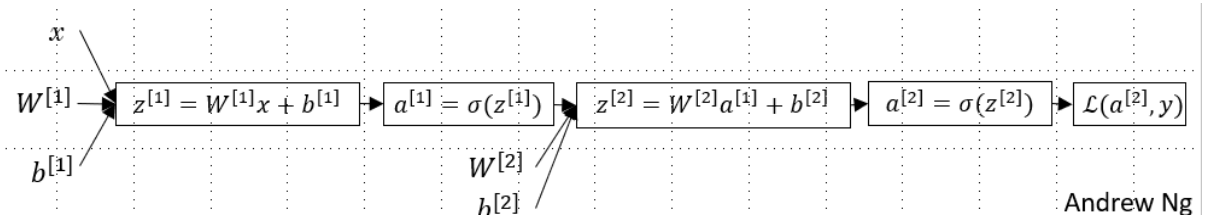
- **Aprendizaje profundo (Deep learning)_ redes neuronales: regresión y clasificación (predicción)**
  - ***Por ejemplo, una clasificación Logit es una red de una neurona***





  - ***Red neuronal: una red con varias capas***
    - En el medio se encuentra la **capas ocultas** que se **calculan automáticamente**.
    - Las siguientes capas se calculan utilizando como entrada la salida de la capa anterior



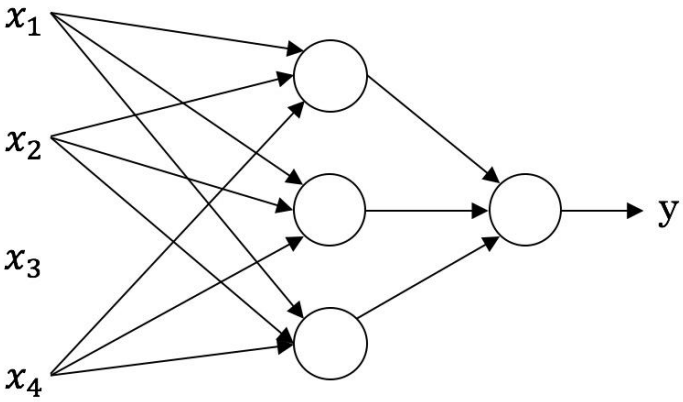


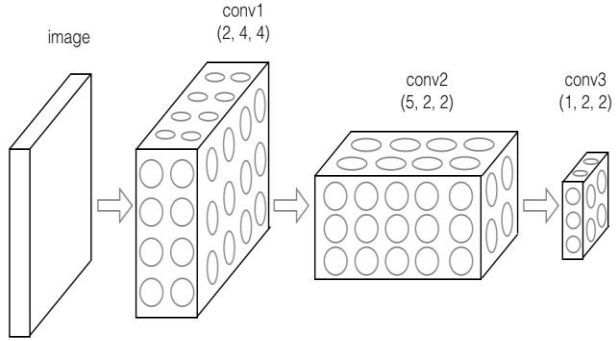
Andrew Ng_deeplearning.ie
elaboración propia

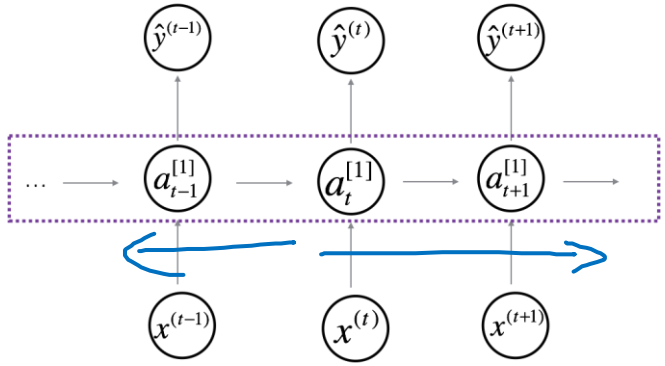
Redes Neuronales: Uso de aprendizaje supervisado

De entrada (x)	De salida (y)	Solicitud
características de la casa	Precio	Bienes raíces
		Publicidad online
Imagen	Objeto (1,..., 1000)	el etiquetado de fotos
Audio	transcripción del texto	Reconocimiento de voz
Inglés	chino	Máquina traductora
La imagen, la información de radar	Posición de otros vehículos	conducción autónoma

Ejemplos de Redes Neuronales: aplicaciones supervisadas y no supervisadas







NN estándar

convolucional NN

recurrente NN

otras perspectivas

- **Hay muchas otras técnicas y algunas otras categorías para fines específicos como los sistemas de recomendación:**
  - hay una categoría mixta, ya que utiliza algunas de las técnicas anteriores tipos:
    - Filtración colaborativa
      - Usuario-Usuario
      - Punto-Punto
      - Reducción de dimensionalidad
    - Motores de búsqueda

- **También un enfoque completamente diferente de otro son Redes**
  - Las redes que representan relaciones subyacente reales (sociales, económicos o de cualquier otro tipo)
  - No son diferentes a las redes computacionales
  - *Google Search utiliza como parte de su motor de este enfoque: Pagerank*

Libros LECTURAS aplicado a esta PRESENTACIÓN

Libro2: Introducción al aprendizaje estadístico (James, Witten)

- Capitulo 2

Book3: Introducción a la recuperación de la información (Manning, Raghavan, Schüzte)

- capítulo 13