Curso de Inteligencia Artificial

Introducción al Aprendizaje Automático

Gibran Fuentes Pineda
IIMAS, UNAM

Definición

- El aprendizaje automático es el estudio de los metodos para programar las computadoras para que aprendan

Thomas G. Dietterich

- Un programa aprende de la experiencia *E* con respecto a la tarea *T* y una medida de rendimiento *P*, si el rendimiento en *T* medido por *P* mejora con *E*

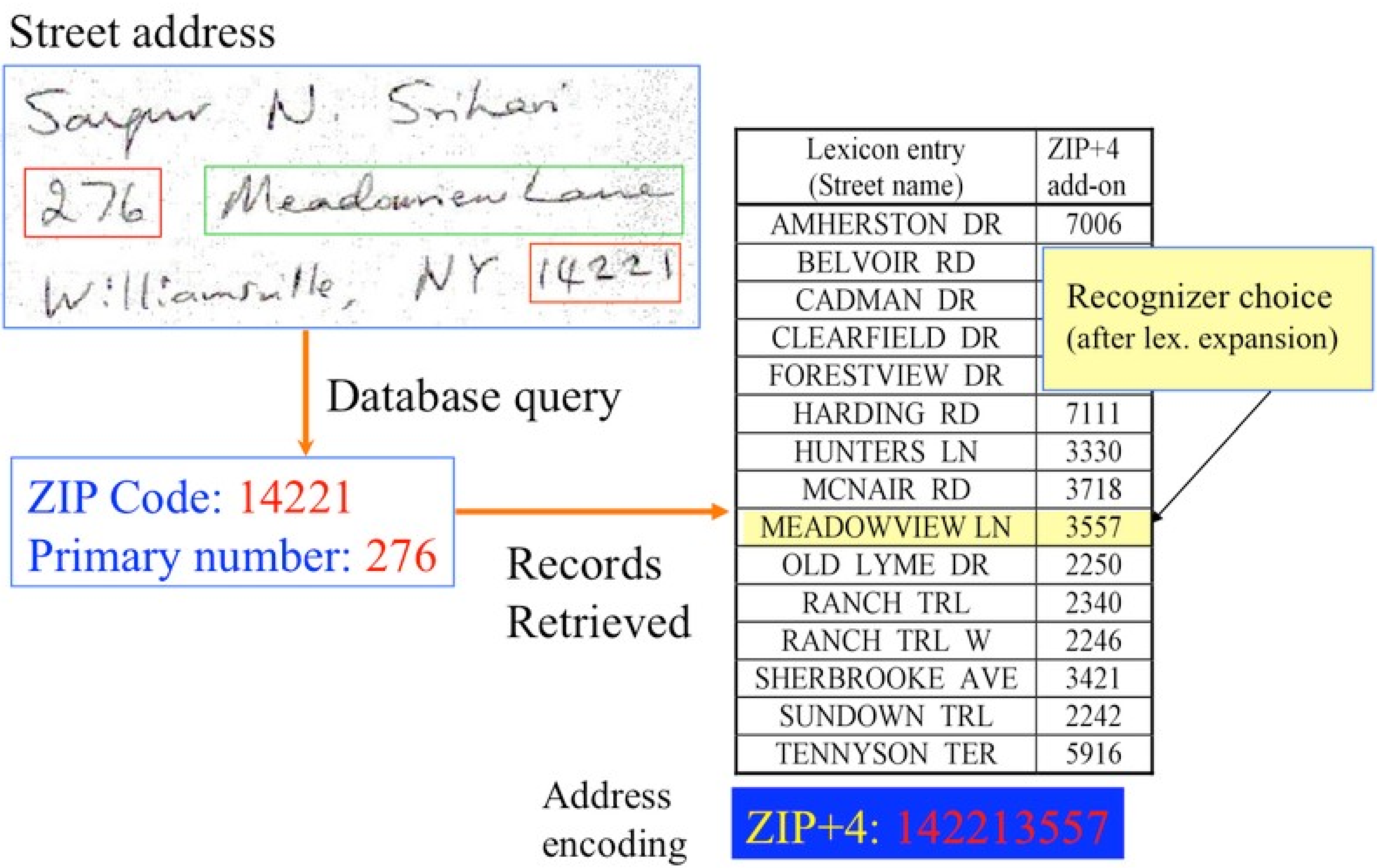
Tom M. Mitchell, Machine Learning

Aplicaciones

- Aprender a manejar un auto
  - *T*: manejar en carreteras usando visión por computadora
  - *P*: promedio de distancia antes de que ocurra un error
  - *E*: imágenes y comandos de manejo observadas en un conductor humano

Aplicaciones

- Reconocimiento de escritura



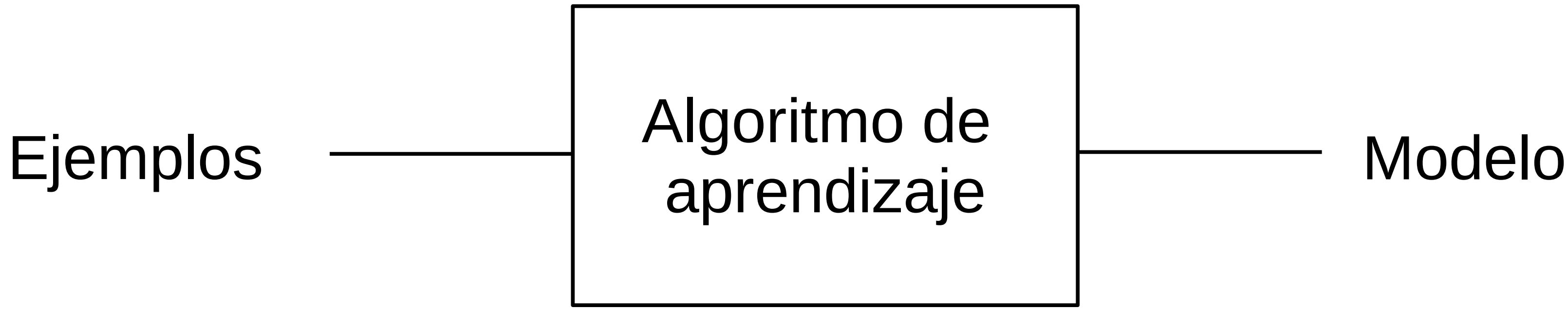
Lexicon entry (Street name)	ZIP+4 add-on
AMHERSTON DR	7006
BELVOIR RD	
CADMAN DR	
CLEARFIELD DR	
FORESTVIEW DR	
HARDING RD	7111
HUNTERS LN	3330
MCNAIR RD	3718
MEADOWVIEW LN	3557
OLD LYME DR	2250
RANCH TRL	2340
RANCH TRL W	2246
SHERBROOKE AVE	3421
SUNDOWN TRL	2242
TENNYSON TER	5916

Tipos de Aprendizaje

- Supervisado
  - Todos los datos de entrenamiento tienen sus correspondientes salidas deseadas
- Semi-supervisado
  - Contiene sólo algunas salidas deseadas
- No supervisado
  - No incluyen salidas deseadas
- Por Refuerzo
  - No hay datos de entrenamiento
  - Busca acciones que maximicen la recompensa

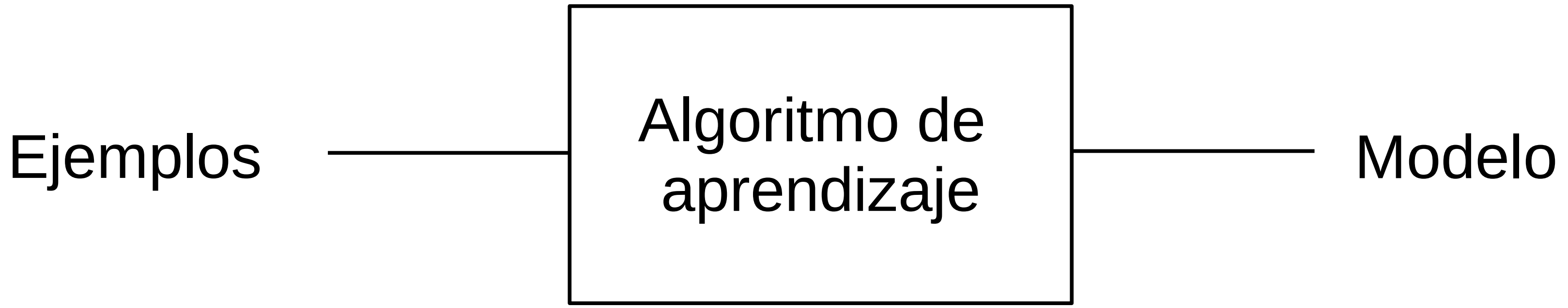
Esquema General

Fase de Entrenamiento



Esquema General

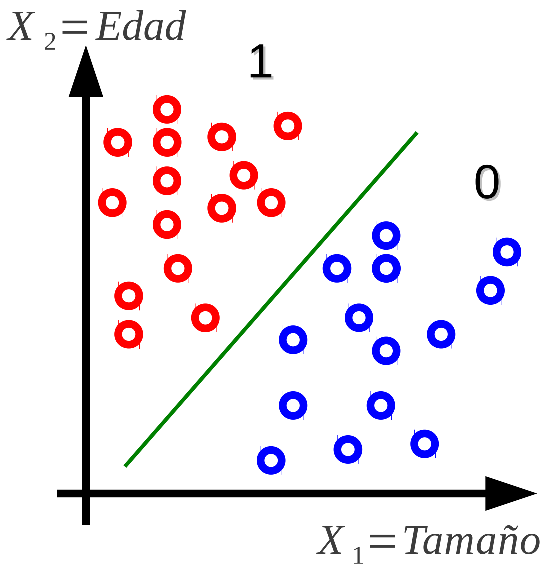
Fase de Entrenamiento



Fase de Prueba

Clasificación y Regresión

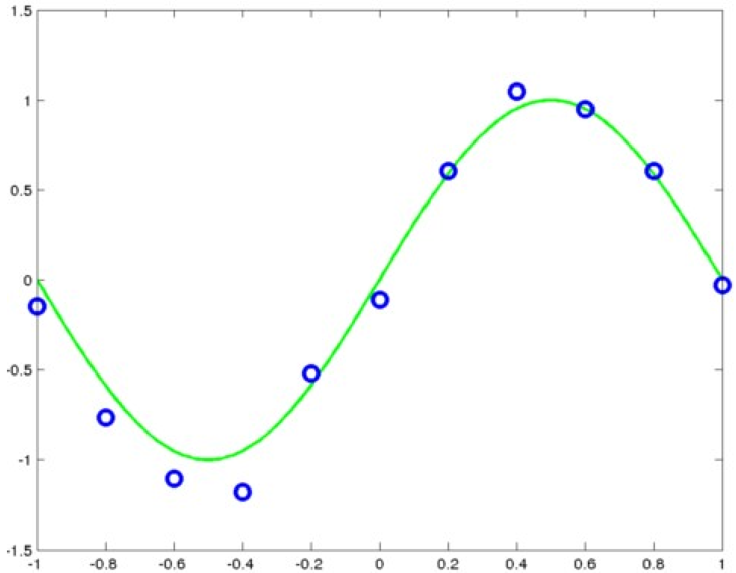
- Clasificación
  - Salida discreta finita
  - Redes neuronales, etc



Clasificación y Regresión

- Clasificación
  - Salida discreta finita
  - Redes neuronales, etc

- Regresión
  - Salida continua o infinita
  - Regresión lineal, etc

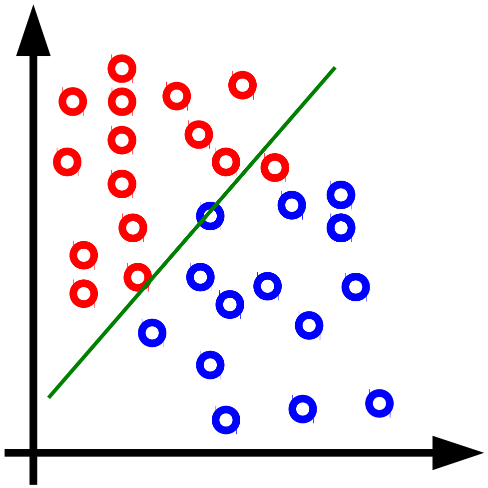


Diseño

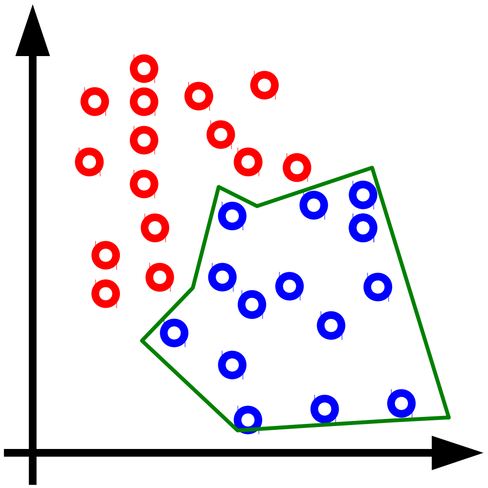
- Descripción del problema
- Recopilación de datos de entrenamiento
- Preprocesamiento de los datos de entrenamiento
- Codificación de datos de entrenamiento
- Elección de modelo de aprendizaje
- Codificación de salidas
- Optimización

Overfitting

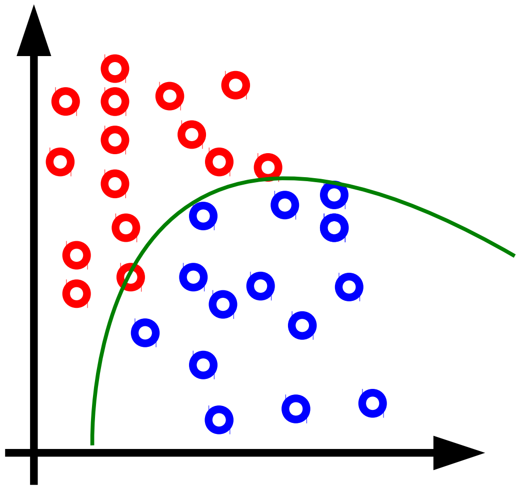
Underfitting



Overfitting



Fair Fitting



La Maldición de la Dimensionalidad

- Crecimiento exponencial en el número de datos requeridos para muestrar un espacio con una precisión dada
- Overfitting
- Complejidad computacional

Reducción de Dimensiones

- Evitar características no necesarias
  - No discriminantes
  - Dependientes
- Selección de características: Elegir un subconjunto de características distintivas
  - Fisher Linear Discriminant
  - Principal Component Analysis
- Extracción de características: Crear nuevas características a partir de las existentes

Incertidumbre

- En muchos problemas no es posible determinar la situación completa del entorno
- Prescencia de ruido
- No es posible predecir lo que pasará al realizar una acción
- La complejidad de algunos problemas puede ser extremadamente grande

La Pregunta

- ¿Lloverá mañana?
  - Basado en frecuencias
  - Basado en evidencias
- ¿Quién ganará el campeonato?
  - Toluca o América o Pumas

Teoría de Probabilidades

- Es una forma matemática de describir la incertidumbre
- Dos elementos principales:
  - Espacio de muestreo Ω
  - Ley de la probabilidad

Teoría de Probabilidades

- Existe una probabilidad del 90% que el medicamento X funcione
  - En 9 de 10 casos funciona
- Existe una probabilidad del 90% que la Iliada y la Odisea fueron escritos por el mismo autor
  - Una creencia

Espacio de Muestreo

- Todos los posibles resultados de un experimento exhaustivamente
  - Lanzar una moneda al aire $\Omega=\{A,S\}$
  - Lanzar un dado $\Omega=\{1,2,3,4,5,6\}$
  - Lanzar tres moneda al aire

    $\Omega=\{AAA, AAS, ASA, SAA, ASS, SAS, SSA, SSS\}$
  - La primera palabra de una oración

    $\Omega=\{A, ábaco, Ana, azar, \ldots\}$
  - Que una persona diga una oración

    $\Omega=\{S|S\, sea\, una\, oración\, válida\}$

Experimento

- Un proceso con varios posibles resultados
  - Muy similar a un máquina no determinista
- Tipos de experimentos
  - Juegos de azar
  - Procesos económicos o políticos
  - Efectos de medicamentos
- Al resultado ocurrido se le conoce como evento

Algunas Propiedades

- $0 \leq P(A) \leq 1$
- $P(cierto) = 1, P(falso) = 0$ $P(cierto) = 1, P(falso) = 0$
- $Si\ A\ es\ un\ subconjunto\ de\ B\ entonces\ P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A\ U\ B) = P(A) + P(B)\ si\ A\ y\ B\ son\ disjuntos$
- $P(\Omega) = 1$
- $P(A \cup B) \leq P(A) + P(B)$
- $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$

Divide y Vencerás

- La probabilidad de un evento complejo S es la suma de las probabilidades de sus elementos

$$P(\{A_{1,} A_{2,} \ldots, A_n\}) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

El Juego

- Definir un modelo, es decir, un espacio de muestreo y sus probabilidades

- Dado un modelo derivar probabilidades de ciertos eventos

Probabilidad Condicional

- Nos permite razonar con información parcial
  - Lanzamos 3 monedas al aire, sabemos que la segunda y tercera fueron SOL cual es la probabilidad que la primera sea AGUILA

$$P\left(A_1 X_2 X_3 \middle| X_1 S_2 S_3\right)$$

  - Probabilidad que una persona tenga caries si sabemos que le duele la muela

$$P\left(caries \middle| dolor\, de\, muelas\right)$$

Enfoque Frecuentista

- Lanzar tres monedas al aire 5 veces

$AAS\ ,ASS\ ,SAA\ ,ASA\ ,ASS$

- ¿Cuál es la probabilidad que al menos dos monedas salgan sol?

Enfoque Frecuentista

- Lanzar tres monedas al aire 5 veces

$$\boxed{AAS}, ASS, \boxed{SAA}, \boxed{ASA}, ASS$$

- ¿Cuál es la probabilidad que al menos dos monedas salgan sol?

$$P(al\ menos\ dos\ soles) = 3/5$$

Enfoque Frecuentista

- Lanzar tres monedas al aire 5 veces

$$AAS\,,ASS\,,SAA\,,ASA\,,ASS$$

- ¿Cuál es la probabilidad que al menos dos monedas salgan sol?

$$P\left(al\,menos\,dos\,soles\right)=3/5$$

- ¿Cuál es la probabilidad de que la primera sea águila?

Enfoque Frecuentista

- Lanzar tres monedas al aire 5 veces

$$AAS\ ,\ ASS\ ,\boxed{SAA},\ ASA\ ,\ ASS$$

- ¿Cuál es la probabilidad que al menos dos monedas salgan sol?

$$P\left(al\ menos\ dos\ soles\right)=3/5$$

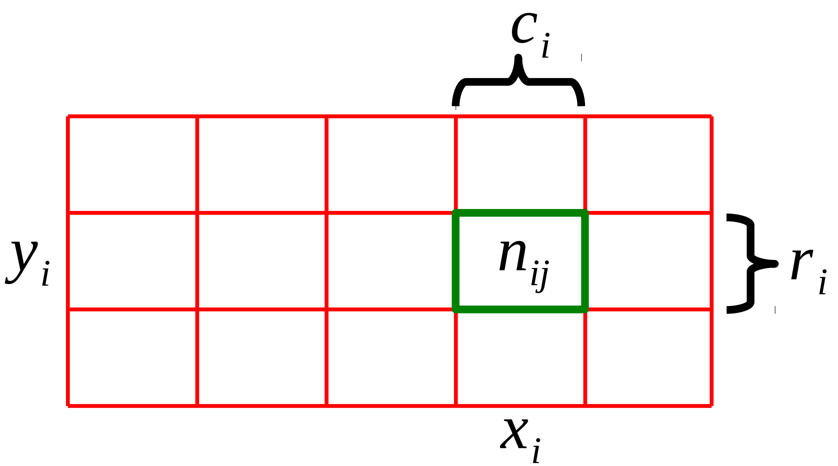
- ¿Cuál es la probabilidad de que la primera sea águila?

$$P\left(primera\ águila\right)=1/5$$

La Regla de la Suma

- Probabilidad que X tome valor x_i y Y tome valor y_j

$$P(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$



La Regla de la Suma

- Probabilidad que X tome valor x_i y Y tome valor y_j

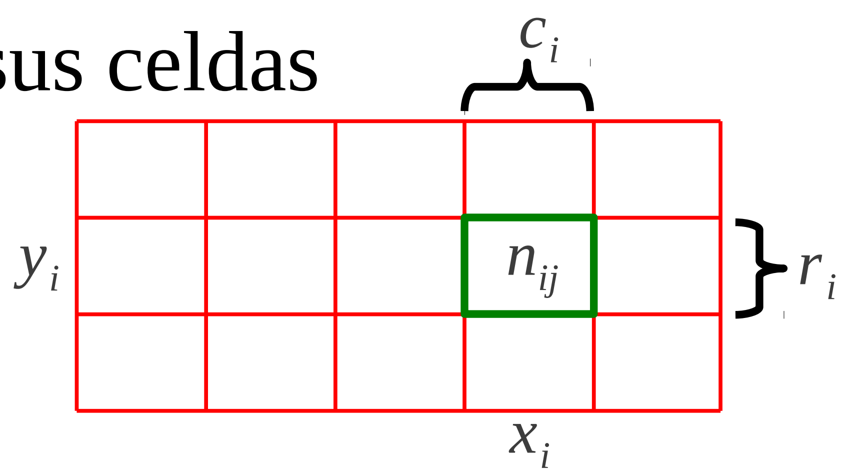
$$P(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

- Probabilidad que X tome valor x_i sin importar el valor de Y

$$P(X = x_i) = \frac{c_i}{N}$$

- Como la probabilidad de c_i es la suma de las probabilidades de cada una de sus celdas

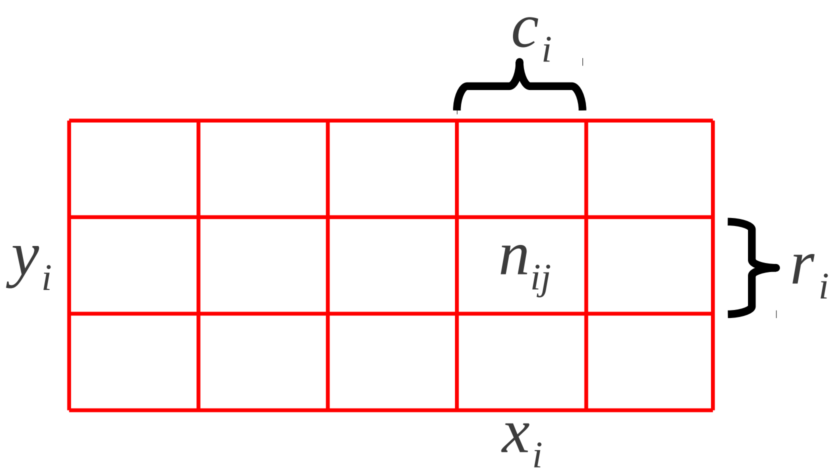
$$P(X = x_i) = \sum_{j=1}^{L} P(X = x_i, Y = y_i)$$



La Regla del Producto

- La probabilidad condicional se puede obtener encontrando la fracción de elementos de la columna c_i que caen en la celda i,j.

$$P\left(Y=y_i \middle| X=x_i\right)=\frac{n_{ij}}{c_i}$$



$$P\left(X=x_i, Y=y_j\right)=\frac{n_{ij}}{N}=\frac{n_{ij}}{c_i}\frac{c_i}{N}$$

$$P\left(X=x_i, Y=y_j\right)=P\left(Y=y_i \middle| X=x_i\right)P\left(X=x_i\right)$$

Regla de Bayes

- Relaciona las probabilidades en una proposición antes y después considerar **evidencia**

A Posteriori
Probabilidad de la hipótesis dada cierta evidencia

Verosimilitud
Probabilidad de la evidencia dada la hipótesis

A Priori
Probabilidad de la hipótesis antes de la evidencia

$$\boxed{P(A|B)} = \frac{\boxed{P(B|A)}\boxed{P(A)}}{\boxed{P(B)}}$$

Marginal
Probabilidad de la evidencia bajo todas las posibles hipótesis

Regla de Bayes: Intuición

- Relaciones causa-efecto
  - Un mamografía muestra una mancha
  - La mancha es el efecto o la evidencia, la causa puede ser un tumor maligno, un tumor no maligno u otra causa
- Durante el diagnóstico vemos la mancha en la mamografía y nos gustaría saber la probabilidad de que sea un tumor

Regla de Bayes: Intuición

$$Causas = \{A_{1,} A_{2,} A_3\}$$

$$Evidencia = B$$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

El Problema Según Bayes

$$P(C|M_p)=\frac{P(M_p|C)\,P(C)}{P(M_p)}$$

El Problema Según Bayes

$$P(C|M_p)=\frac{P(M_p|C)P(C)}{P(M_p)}$$

$$P(M_p)?$$

El Problema Según Bayes

$$P(C|M_p)=\frac{P(M_p|C)P(C)}{P(M_p)}$$

$$P(M_p)=P(M_p\cap C)+P(M_p\cap nC)$$

$$P(M_p)=P(M_p|C)P(C)+P(M_p|nC)P(nC)$$

$$P(M_p)=0.8\times 0.01+0.096\times 0.99=0.103$$

$$P(C|M_p)=\frac{0.8\times 0.01}{0.103}=0.0769$$

El Problema

- El contexto
  - 1% de las mujeres mayores de 40 años que se checan regularmente tiene cancer
  - 80% de estás mujeres recibirán el resultado de una mamografía positiva
  - 9.6% de la mujeres sin cancer también recibirán una mamografía positiva
- Una mujer de 45 años recibe positivo el resultado de una mamografía. ¿Cúal es la probabilidad de que tenga cancer?

Los Hechos

- Alrededor de 15% de los doctores razonan adecuadamente este problema.
- Un cambio en la escala puede elevar a 46% la cantidad de doctores que razonan adecuadamente el problema.

Revisando Datos

$$P(C)=0.01$$

$$P(Mp|C)=0.8$$

$$P(Mp|nC)=0.096$$

Independencia

- Dos eventos son independientes cuando:

$$P(A|B)=P(A)$$

- Lo que lleva a

$$P(A\cap B)=P(A)P(B)$$

- Fácil! Si *A* es producido por una máquina *1* y *B* por otra máquina *2*, son independientes

Independencia: Concepciones Erróneas

- En términos del espacio de muestreo es un error pensar que *A* y *B* son independientes si son disjuntos.

- Lanzo dos monedas al aire ¿es la primera moneda independiente de la segunda?
  - Ω={AS,SA,SS,SA}
  - $P(X1 \cap X2)=1/4$
  - $P(X1) = 1/2$ $P(X2)=1/2$

Asumir Independencia

- Asumir independencia ayuda a simplificar los cálculos
- Muchos modelos probabilísticos usados en aprendizaje automático asumen independencia

Variable Aleatoria

- Hasta ahora la función P ha sido definida para aceptar conjuntos de Ω
  - Idea: estandarizar el argumento de P

- Un variable aleatoria hace este trabajo, es una función de Ω a R

Variable Aleatoria

- Todos los posibles resultados de un experimento exhaustivamente
  - Lanzar una moneda al aire

$$\Omega=\{A,S\},X=\{1,0\}$$

  - Lanzar un dado

$$\Omega=\{1,2,3,4,5,6\},X=\{1,2,3,4,5,6\}$$

  - Lanzar tres moneda al aire

$$\Omega=\{AAA,AAS,ASA,SAA,ASS,SAS,SSA,SSS\}$$

$$X=\{0,1,2,3\}$$

Modelos Discriminativos vs Generativos

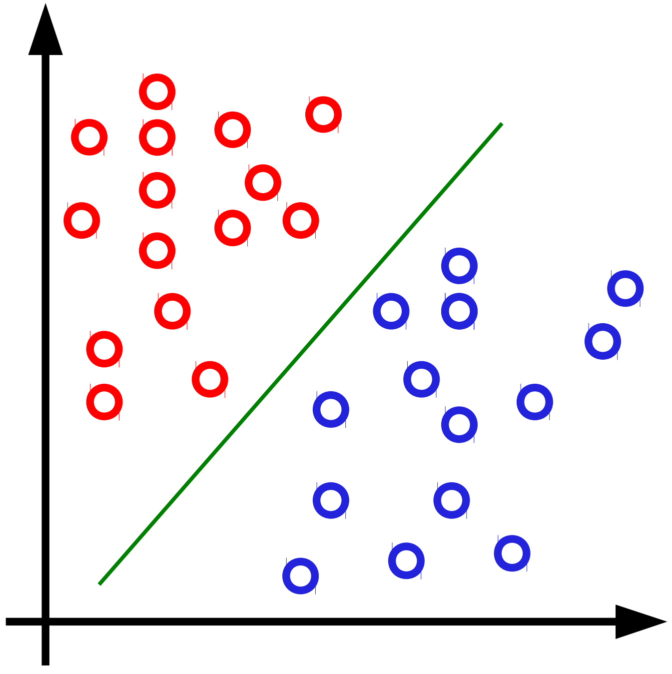
- Problema: Clasificar frutas
- Estrategia Generativa
  - Modelar cada clase de fruta para determinar cuál es
- Estrategia Discriminativa
  - Determinar las diferencias entre las clases de frutas

Modelos Discriminativos

- Entrenamiento: Estiman probabilidades a posteriori
  - No intentan modelar los datos
  - Definen fronteras de decisión
- Reconocimiento: Predicen la clase a partir de las fronteras de decisión
- Ejemplos: SVMs, Neural Networks, etc.

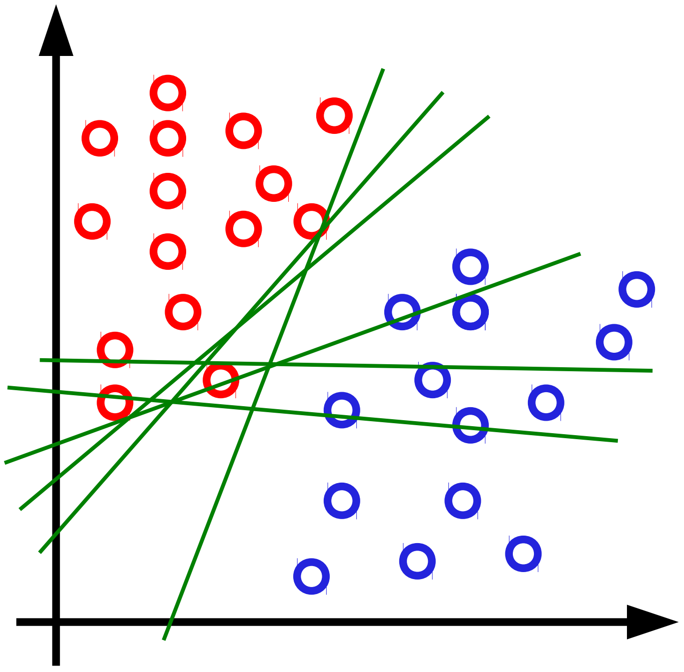
SVM Lineal

- Puede verse como la separación de clases en el espacio de características por un hiperplano



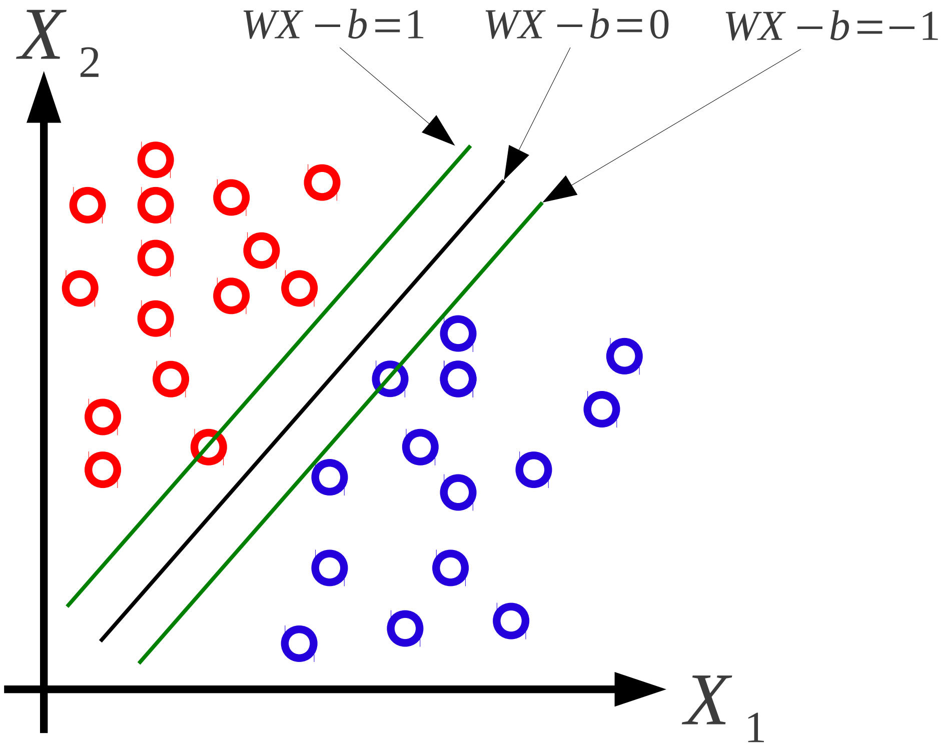
¿Qué Hiperplano Elegir?

- Muchas posibles soluciones



¿Qué Hiperplano Elegir?

- Muchas posibles soluciones
- Utiliza los vectores de soporte que maximizan el margen



Modelos Generativos

- Entrenamiento: Modelan distribuciones conjuntas sobre todas las entradas y salidas
  - Modelo aproximado de cada clase
  - Es posible “generar” datos sintéticos muestreando la distribución conjunta de probabilidad
- Reconocimiento: Asignan la clase con mayor probabilidad
- Ejemplos: Naive Bayes, Modelos Ocultos de Markov, etc.

Naive Bayes

- Usar la regla de Bayes para clasificación

Qué tan probable es la característica en la clase

Qué tan probable es la clase

Probabilidad de la clase dada cierta característica

$$\boxed{P(C_i|X)} = \frac{\boxed{P(X|C_i)}\boxed{P(C_i)}}{\boxed{P(X)}}$$

Qué tan probable es la característica

Ejemplo

- Determinar si una persona es mujer u hombre dado su nombre

$$P(mujer|maria) = \frac{P(maria|mujer)P(mujer)}{P(maria)}$$

$$P(hombre|maria) = \frac{P(maria|hombre)P(hombre)}{P(maria)}$$

Nombre	Sexo
María	**Mujer**
Pedro	Hombre
María	Hombre
Inés	**Mujer**
Aurora	Mujer
Roberto	Hombre
María	**Mujer**
Elsa	Mujer

Más Características

- Presuposiciones:
  - Todas la características son independientes dada la clase
  - Todas las características son igualmente importantes
- Bayes con *n* características

$$P(C_i|X_1,\ldots X_n)=\frac{P(X_1|C_i)\times\ldots\times P(X_n|C_i)}{P(X_1)\times\ldots\times P(X_n)}P(C_i)$$