UNIVERSIDAD ADOLFO IBÁÑEZ

S02.1: Clasificación Lineal

Aprendizaje

Dr. Juan Bekios Calfa

Magíster en *Data Science*
Facultad de Ingeniería y Ciencias

Información de Contacto

- Juan Bekios Calfa
    - email: juan.bekios@edu.uai.cl
    - Web page: http://jbekios.ucn.cl
    - Teléfono: 235(5162) - 235(5125)

Contenidos

Contenidos

Tareas de clasificación

Tareas de regresión: predicción de cantidad real $y \in \mathbb{R}$

Tareas de clasificación: predicción de *cantidades discretas* y

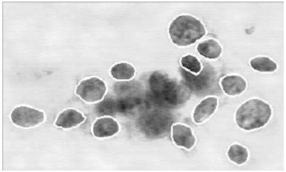
Clasificación binaria: $y \in \{-1,\ +1\}$

Clasificación multiclase: $y \in \{1, 2, \cdots,\ k\}$

Ejemplo: clasificación cáncer de mama

Un ejemplo de clasificación conocido utilizando el aprendizaje automático: Diagnosticar si un el tumor de mama es benigno o maligno [Street et al., 1992]

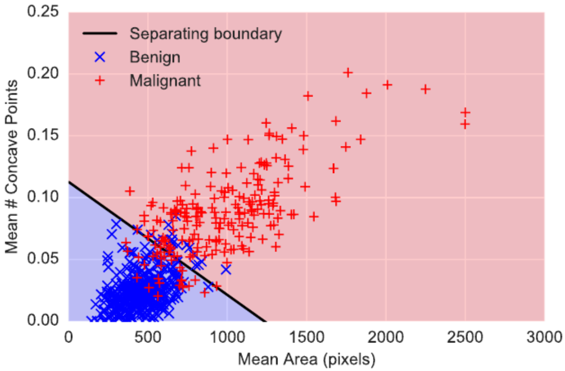
Contexto: el médico extrae una muestra de líquido del tumor, tiñe las células y luego describe varias de las células (el procesamiento de imágenes refina el contorno)



El sistema calcula características para cada célula, como área, perímetro, concavidad, textura (10 en total); calcula la media / estándar / máxima para todas las características

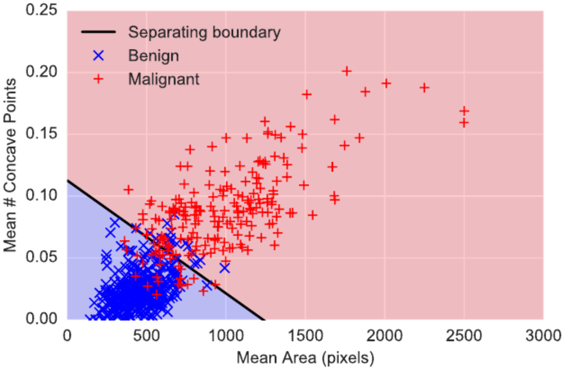
Ejemplo: Clasificación del cáncer de mama

Trazado de dos características: área media vs. puntos cóncavos medios, para dos clases



Ejemplo: Clasificación lineal

Clasificación lineal ≡ “dibujar lineas para la separación de las clases”



Contenidos

Descripción formal

Características de entrada: $x^{(i)} \in \mathbb{R}^n, i = 1,, m$

$$Ej : x^{(i)} = \begin{bmatrix} area\_media^{(i)} \\ puntos\_concavos\_medios^{(i)} \\ 1 \end{bmatrix}$$

Salidas:

$$y^{(i)} \in H, i = 1, ..., m$$

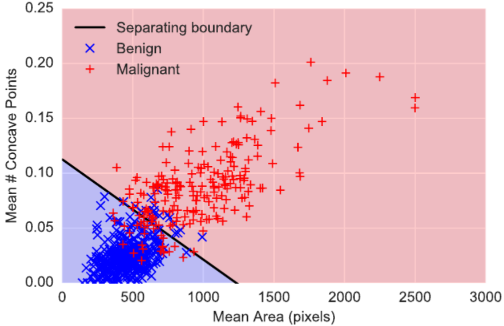
$$Ej : y^{(i)} \in \{-1(benigno), +1(maligno)\}$$

Parámetros del modelo: $\theta \in \mathbb{R}^n$

Función de hipótesis: $h_\theta : \mathbb{R}^n \to \mathbb{R}$, apunta al mismo signo que la salida (se puede decir informalmente, que es una medida de confianza en nuestra predicción)

Ej: $h_\theta(x) = \theta T x, \;\; \hat{y} =$sign $(h_\theta(x))$

Entendiendo los diagramas de clasificación lineal



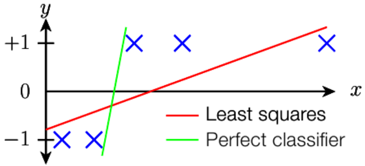
El color muestra regiones donde $h_\theta(x))$ es positivo

El límite de separación está dado por la ecuación $h_\theta(x)) = 0$

Funciones de pérdida para clasificación

¿Cómo definimos una función de pérdida $\ell : \mathbb{R} \times \{-1, +1\} \to \mathbb{R}_+$?

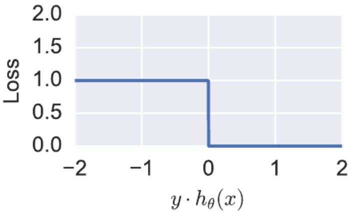
¿Qué pasa con solo usar la pérdida al cuadrado?



Función de pérdida 0/1

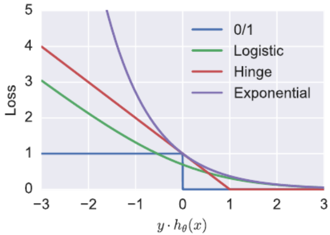
Queremos minimizar la siguiente función de perdida (llamada pérdida 0/1, o simplemente ”error”)

$$
\begin{aligned}
\ell_{0/1}(h_\theta(x), y) &= \begin{cases} 0 \; si \; sign(h_\theta(x)) = y \\ 1 \; en \; caso \; contrario \end{cases} \\
&= 1\{y \cdot h_\theta(x) \leq 0\}
\end{aligned}
$$



Otras funciones de pérdida

Lamentablemente, la pérdida 0/1 es difícil de optimizar (Es NP-hard encontrar un clasificador con una perdida 0/1 mínima, relacionado a una propiedad llamada convexidad de la función)
En su lugar, se utilizan varias pérdidas alternativas para la clasificación.



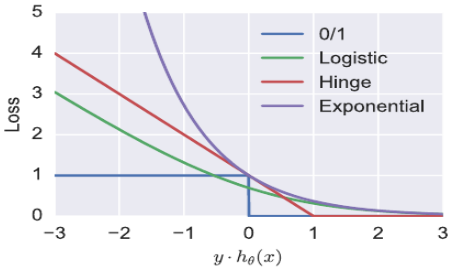
$$\ell_{0/1} = 1\{y \cdot h_\theta(x) \leq 0\}$$
$$\ell_{\text{logistic}} = \log(1 + \exp(-y \cdot h_\theta(x)))$$
$$\ell_{\text{hinge}} = \max\{1 - y \cdot h_\theta(x), 0\}$$
$$\ell_{\exp} = \exp(-y \cdot h_\theta(x))$$

13

Pregunta: Sensibilidad a *outliers*

¿Qué tan sensible estimarías que sería cada una de las siguientes pérdidas para valores atípicos o *outliers* (es decir, puntos típicamente muy mal clasificados)?

1. 0/1 < Exp < {Hinge,Logitistic}
2. Exp < Hinge < Logistic < 0/1
3. Hinge < 0/1 < Logistic < Exp
4. 0/1 < {Hinge,Logistic} < Exp



5. Los valores atípicos no existen en la clasificación porque el espacio de salida está limitado

Optimización de aprendizaje automático

Con esta notación, el problema de aprendizaje automático ”canónico” se escribe exactamente igual

$$\text{Minimizar} \sum_{i=1}^{m} \ell(h_\theta \theta(x^{(i)}),\ y^{(i)})$$

A diferencia de los mínimos cuadrados, no existe una solución analítica para la condición de gradiente cero para la mayoría de las pérdidas de clasificación

En cambio, resolvemos estos problemas de optimización utilizando el descenso de gradiente (o un método de optimización alternativo, pero aquí solo consideraremos el descenso de gradiente)

$$\text{Repetir: } \theta := \theta - \alpha \sum_{i=1}^{m} \nabla_\theta \ell(h_\theta(x^{(i)}),\ y^{(i)})$$

Contenidos

Máquinas de soporte vectorial

Una máquina de vectores de soporte (lineal) (Support Vector Machine o **SVM**) solo resuelve el aprendizaje automático canónico problema de optimización usando una pérdida de hinge e hipótesis lineal, más un término de regularización.

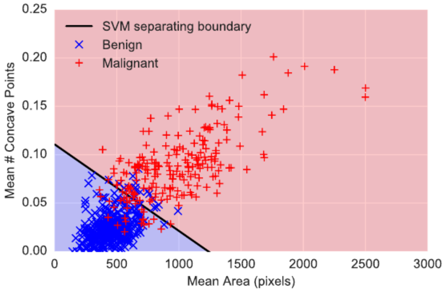
$$\text{Minimizar} \sum_{i=1}^{m} \max\{1 - y^{(i)} \cdot \theta^T x^{(i)},\ 0\} + \frac{\lambda}{2}\|\theta\|_2^2$$

En estricto rigor, el SVM ”estándar” en realidad no regulariza θ_i correspondiente a la característica constante, pero ignoraremos esto aquí
Actualizaciones usando descenso de gradiente :

$$\theta := \theta - \alpha \sum_{i=1}^{m} -y^{(i)} x^{(i)} 1\{y^{(i)} \cdot \theta^T x^{(i)} \leq 1\} - \alpha\lambda\theta$$

Ejemplo de máquina de vectores de soporte

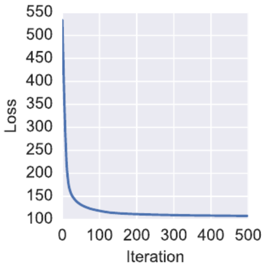
Utilizando la máquina de vectores de soporte en el conjunto de datos de cáncer de mama, con una pequeña regularización parámetro (efectivamente cero)

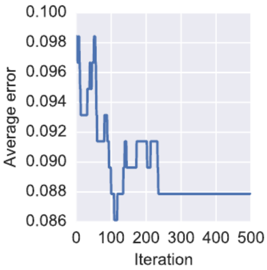


$$\theta = \begin{bmatrix} 1.456 \\ 1.848 \\ -0.189 \end{bmatrix}$$

Progreso de la optimización del SVM

Objetivo de optimización y error versus número de iteración de descenso de gradiente





Contenidos

UNIVERSIDAD ADOLFO IBÁÑEZ

Regresión logística

La regresión logística solo resuelve este problema usando pérdida logística y una hipótesis de función lineal

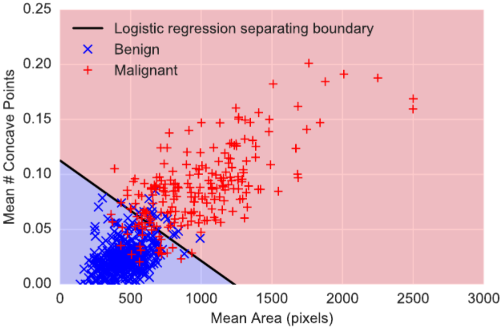
$$\underset{\theta}{Minimizar} \sum_{i=1}^{m} \log(1 + \exp(-y^{(i)} \cdot \theta^T x^{(i)}))$$

Actualizaciones usando descenso de gradiente :

$$\theta := \theta - \alpha \sum_{i=1}^{m} -y^{(i)} x^{(i)} \frac{1}{1 + \exp(y^{(i)} \cdot \theta^T x^{(i)})}$$

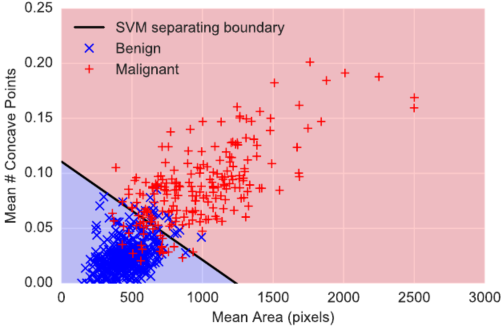
Ejemplo de regresión logística

Utilizando regresión logística en conjunto de datos sobre cáncer, con una pequeña regularización



Ejemplo de regresión logística

Utilizando regresión logística en conjunto de datos sobre cáncer, con una pequeña regularización



Contenidos

Clasificación multiclase

Cuando la salida está en 1, ..., k (por ejemplo, clasificación de dígitos), existen diferentes formas de abordar el problema

Forma 1: Construir k diferentes clasificadores binarios h_{θ_i} con el objetivo de predecir clase i vs todos los demás, las predicciones de salida tienen la forma:

$$\hat{y} = \underset{i}{argmax}\ h_{\theta_{x_i}}(x)$$

Forma 2: Usar una funcion de hipotesis $h_\theta : \mathbb{R}^n \to \mathbb{R}^k$, definir una funcion de perdida alternativa $\ell : \mathbb{R}^k \times \{1,\ ,\ k\} \to \mathbb{R}_+$

Por ejemplo, pérdida de softmax (también llamada pérdida de entropía cruzada):

$$\ell(h_\theta(x),\ y) = \log \sum_{j=1}^{k} \exp(h_\theta(x)_j) - h_\theta(x)_y$$

Referencias

- **Pattern Recognition and Machine Learning.** Christopher M. Bishop. Springer. 2006.
- **The Elements of Statistical Learning: Data Mining, Inference, and Prediction.** Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer. 2016.
- **Practical Data Science: Deep learning**. J. Zico Kolter. `http://www.datasciencecourse.org/slides/deep_learning.pdf`

¿Preguntas?