Regresión Logística

- Se estableció como una técnica en el campo de la estadística a finales de la década de 1940 y la década de 1950
- Técnica de clasificación binaria
  - A pesar de llamarse "regresión"
  - Es un tipo de regresión lineal, pero en lugar de predecir un valor continuo, predice una probabilidad
    - Útil para modelar datos que tienen dos categorías
  - Para ser aplicada a muticlase, estrategia "uno contra todos" o "uno contra uno"

REGRESIÓN LOGÍSTICA

- En la regresión logística, el objetivo es modelar la relación entre las características y la probabilidad de pertenencia a una categoría
  - Por ejemplo, predecir si un email es spam o no spam
  - Determinar la existencia o ausencia de relación entre una o más variables independientes y una variable dependiente dicotómica
    - Es decir, que sólo admite dos categorías que definen opciones o características mutuamente excluyentes u opuestas
  - Estimar o predecir la probabilidad de pertenencia a esa categoría en función de los valores que adoptan las variables independientes

REGRESIÓN LOGÍSTICA

- Si p es la probabilidad de pertenencia a la clase 1
  - 1-p es la probabilidad de pertenencia a la clase 0

$$p(y=0|x) = 1 - p(y=1|x)$$

- El *odds* (razón de probabilidades) es un concepto en estadística que mide la relación entre dos probabilidades
  - El cociente de la probabilidad de pertenecer a una clase y la probabilidad de no pertenecer
    - Es decir, el cociente del número de casos que pertenecen entre el número de casos que no pertenecen

$$odds = \frac{p}{1-p}$$

REGRESIÓN LOGÍSTICA

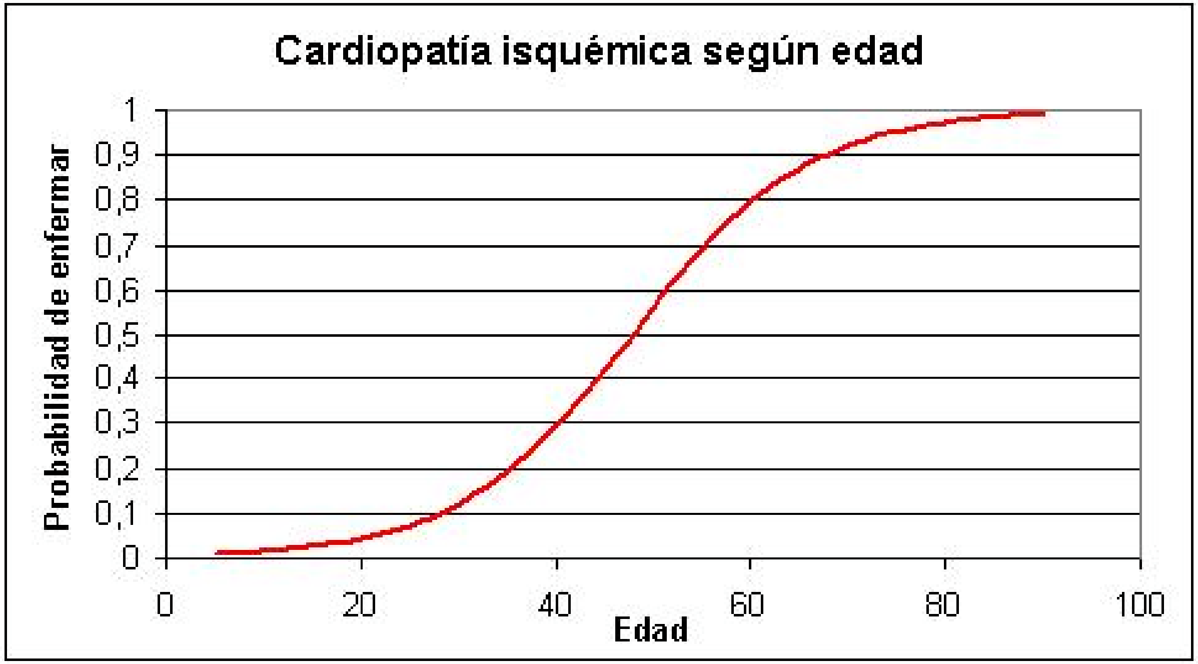
- El *odds* se utiliza comúnmente en problemas de probabilidad y en el análisis de datos
  - Es una manera de expresar cuántas veces más probable es que ocurra un evento en comparación con que no ocurra
  - Proporción entre algo que ocurre y algo que no ocurre
    - Probabilidad: La probabilidad es la relación entre algo que ocurre y todo lo que podría ocurrir.
  - También llamado:
    - Razón de momios
    - Razón de oportunidades
    - Razón de probabilidades

Regresión Logística

- El *odds* toma un valor no negativo
  - Un *odds* de 1 indica que las probabilidades de que ocurra o no ocurra el evento son iguales
  - Un *odds* mayor que 1 indica que el evento es más probable que ocurra
  - Un *odds* menor que 1 indica que el evento es menos probable que ocurra
- El *odds* es especialmente útil cuando se comparan dos resultados, por ejemplo:
  - Apuestas deportivas, estudios médicos, análisis epidemiológicos para evaluar la relación entre exposición y enfermedad, etc.
  - En general, **clasificación binaria**

REGRESIÓN LOGÍSTICA

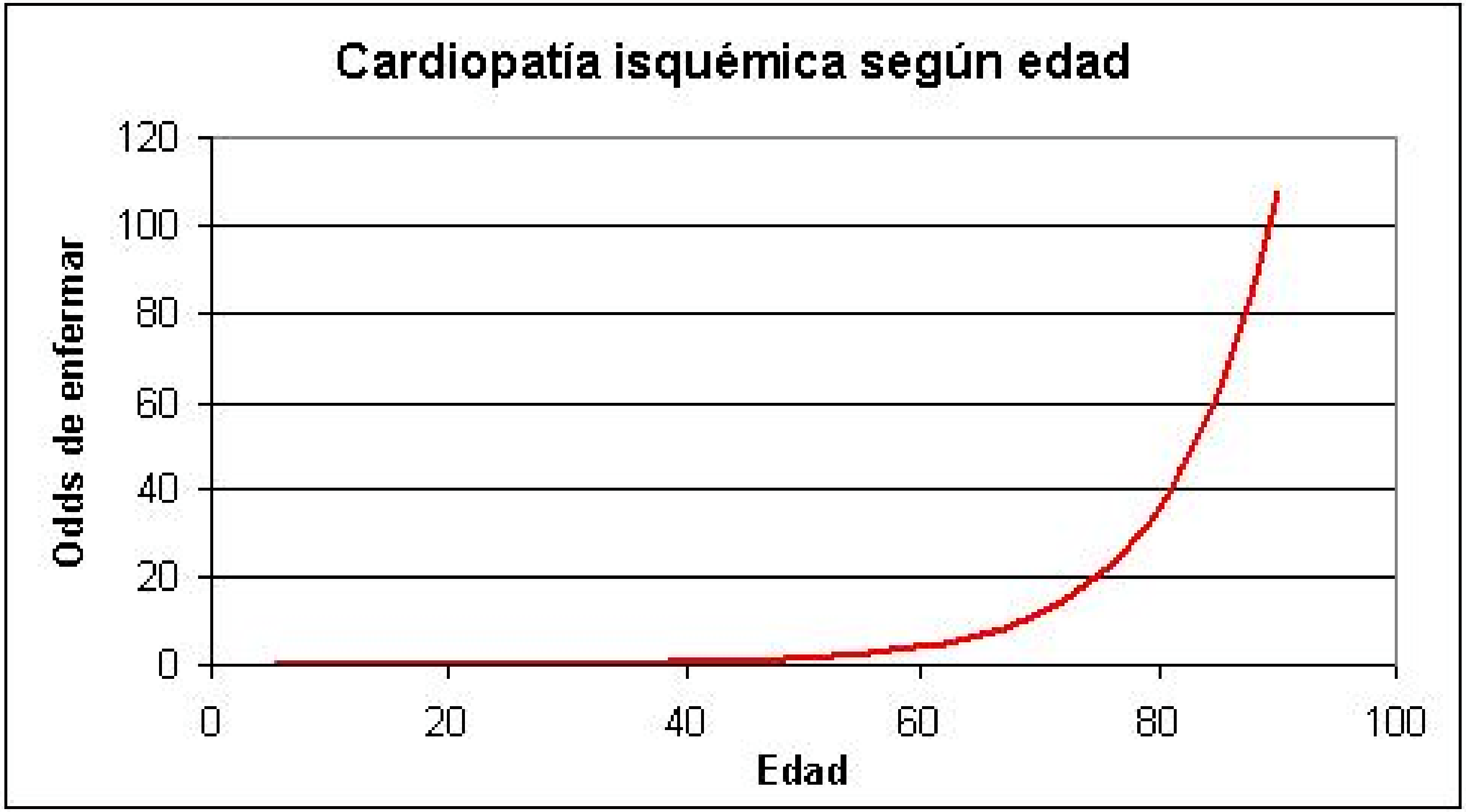
- Ejemplo:
  - Probabilidad de padecer enfermedad coronaria en función de la edad:



  - Relación no lineal, sino con una forma sigmoidea (distribución logística)

REGRESIÓN LOGÍSTICA

- Ejemplo:



REGRESIÓN LOGÍSTICA

- El *log-odds* es el resultado de aplicar un logaritmo natural al *odds*
  - Representa la transformación logarítmica de la relación entre dos probabilidades
  - Logaritmo de la razón de probabilidades

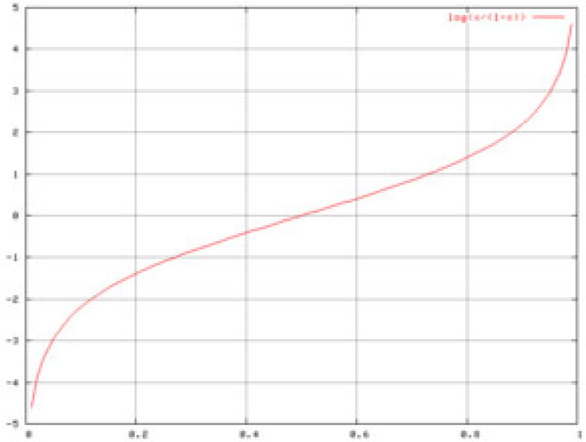
$$log - odds = log(odds) = log\left(\frac{p}{1-p}\right)$$

  - A esta función se llama logit(p)

$$logit(p) = log\left(\frac{p}{1-p}\right) = \mathrm{z}$$

Regresión Logística

- El *log-odds* toma valores en el rango de $-\infty$ a $+\infty$
  - Si p=0.5 → *log-odds*=0
    - p=0.5 → las probabilidades de que ocurra o no ocurra un evento son iguales
  - p se acerca a 1 → el *log-odds* se acerca a ∞
  - p se acerca a 0 → el *log-odds* se acerca a $-\infty$



- Además, la función logit es el negativo de la derivada de la función de entropía binaria

REGRESIÓN LOGÍSTICA

- A partir de esta expresión, se puede despejar p:

$$logit(p) = log\left(\frac{p}{1-p}\right) = z \Rightarrow \frac{p}{1-p} = e^z \Rightarrow p = (1-p)e^z \Rightarrow$$

$$\Rightarrow (1+e^z)p = e^z \Rightarrow p = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

  - Por lo tanto, la probabilidad de un caso positivo se expresa en función de una función **logística**
    - También conocida como función **sigmoide**
  - La regresión logística utiliza la función logística para transformar el *log-odds* en una probabilidad en el rango de 0 a 1
    - Esto permite la modelización y la predicción en problemas de clasificación

REGRESIÓN LOGÍSTICA

- Por tanto, se puede calcular p en función del *log-odds*
- Este valor del *log-odds* será la salida de un modelo cuyas entradas serán las variables independientes
- En regresión logística este será un modelo lineal
  - Es decir, z será una combinación lineal de las características de entrada

$$z = logit(p) = b_0 + \sum b_i x_i$$

Regresión Logística

- ¿Por qué utilizar un modelo lineal?
  - Simplicidad en la modelización:
    - La elección de una relación lineal permite una formulación matemática más simple y computacionalmente eficiente del modelo
    - Los métodos de optimización y ajuste de parámetros también se vuelven más manejables
  - Interpretación:
    - Los cambios en las características se traducen en cambios proporcionales en los *log-odds*
      - Esto es útil para entender cómo las características afectan la probabilidad de pertenencia a una categoría.

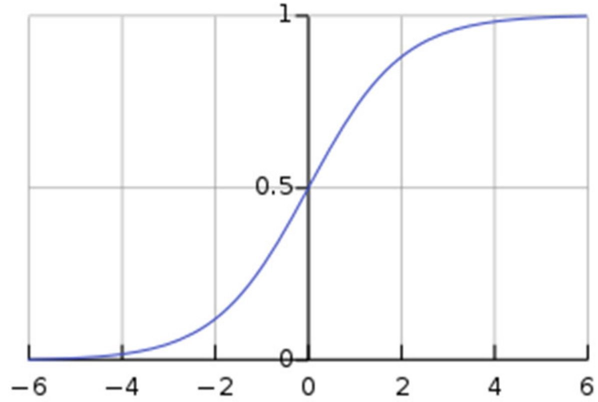
REGRESIÓN LOGÍSTICA

- ¿Por qué utilizar un modelo lineal?
  - Facilita la regularización:
    - Al expresar la relación entre las características y el resultado en forma lineal, es más sencillo aplicar técnicas de regularización como L1 (*Lasso*) o L2 (*Ridge*), que penalizan los coeficientes grandes
      - Esto puede ayudar a prevenir el sobreajuste y mejorar la generalización del modelo
  - Comprobada efectividad empírica:
    - A lo largo de la historia de la estadística y el aprendizaje automático, se ha demostrado que una relación lineal en un espacio logarítmico es una representación efectiva para capturar las relaciones entre características y probabilidades

REGRESIÓN LOGÍSTICA

- La regresión logística combina la función logística con una combinación lineal de las características:
  - Función a utilizar: logística
    - También llamada sigmoide

$$\sigma(z) = \frac{1}{1+e^{-z}}$$



    - Muy usada en redes de neuronas artificiales

  - Valores altos de z → f(z)=1
  - Valores bajos de z → f(z)=0
  - Función continua, y acotada entre 0 y 1
  - z será la combinación lineal de los atributos de entrada x_i

REGRESIÓN LOGÍSTICA

- La regresión logística combina la función logística con una combinación lineal de las características:
  - Modelo de Probabilidad:
    - A través de la función logística (sigmoide), el *log-odds* se transforma en una probabilidad p:

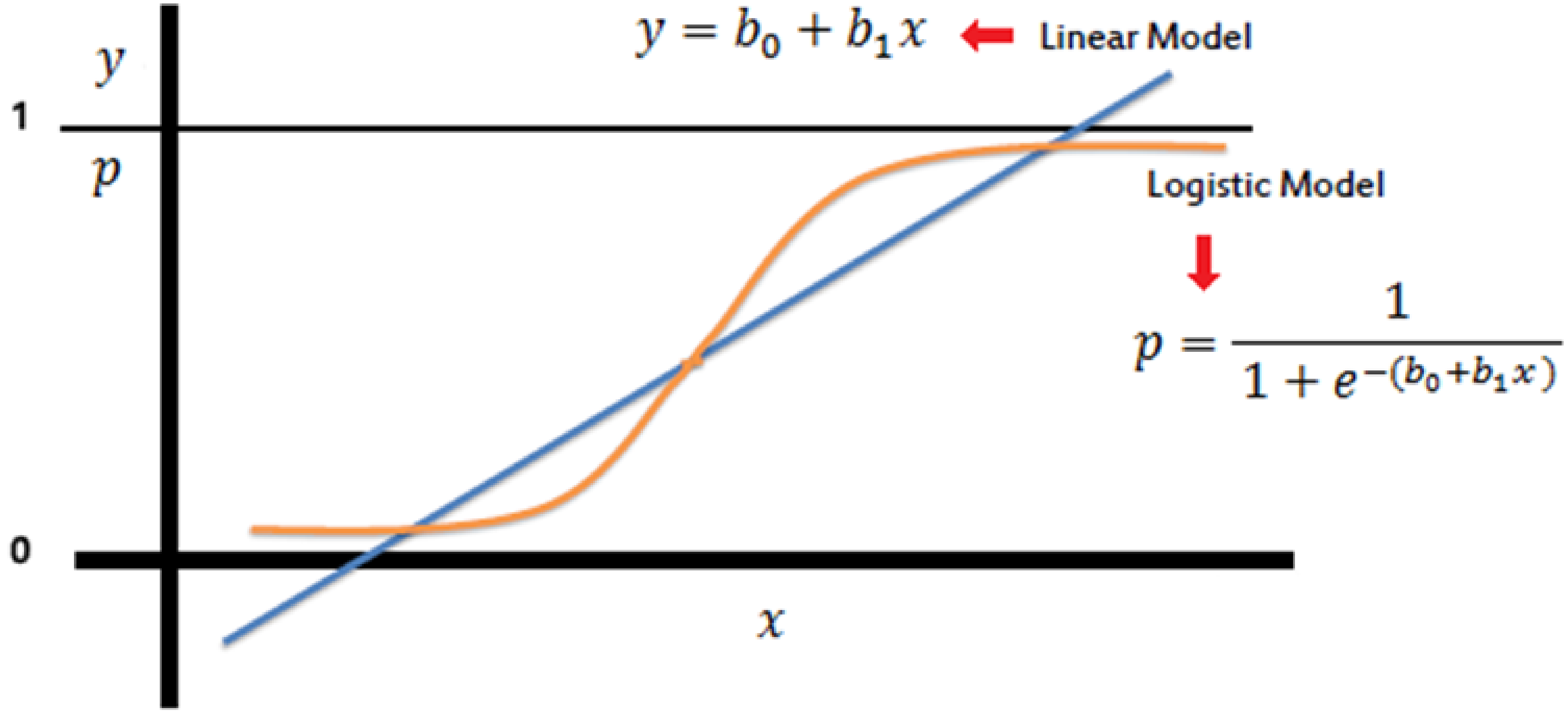
$$p = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-logit(p)}} = \frac{1}{1 + e^{-(b_0 + \sum b_i x_i)}}$$

      - Valor entre 0 y 1

    - Esta es la salida del modelo de regresión logística

REGRESIÓN LOGÍSTICA

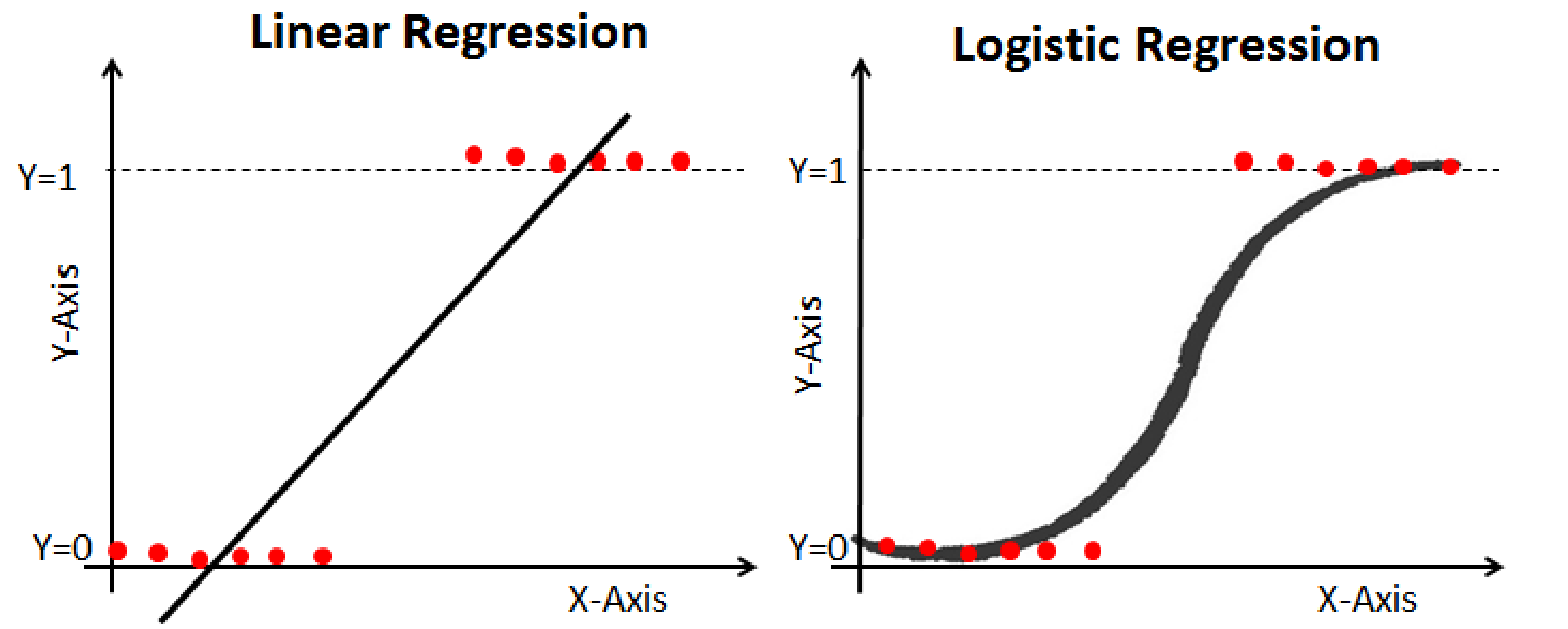
- La regresión logística combina la función logística con una combinación lineal de las características:



- ¿Se podría utilizar la regresión lineal para el cálculo de probabilidades?

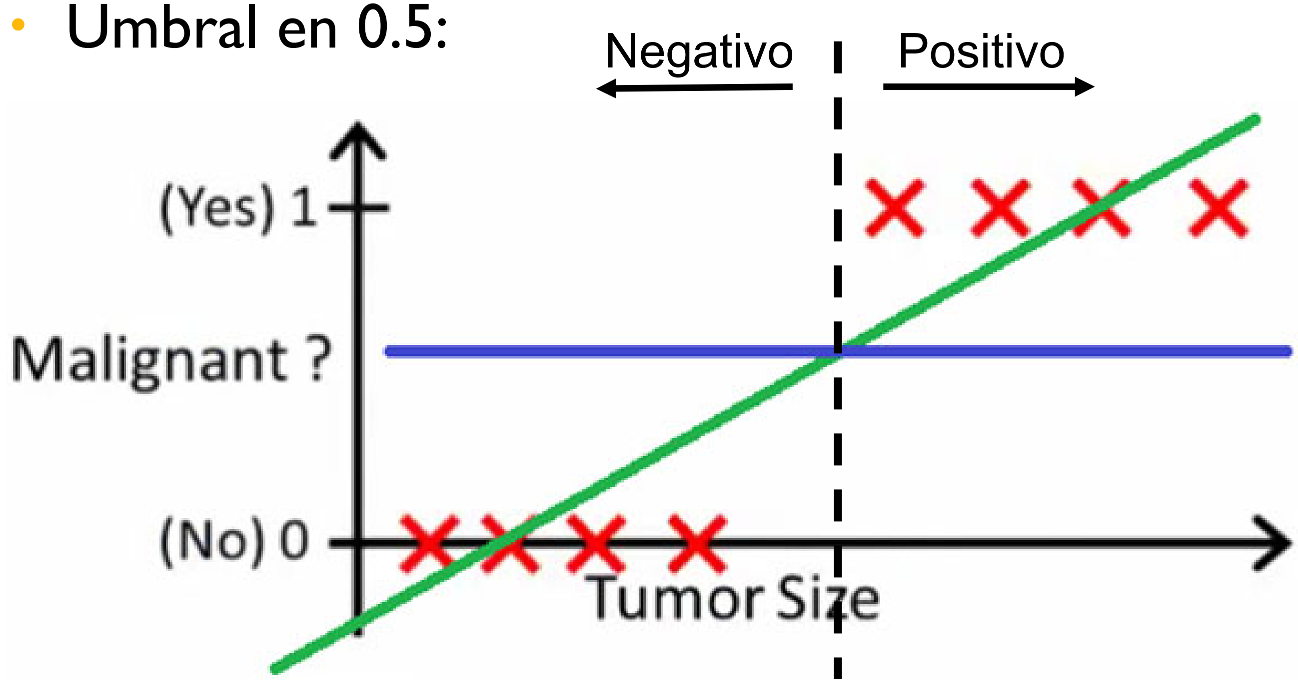
REGRESIÓN LOGÍSTICA

- Regresión lineal vs. Regresión logística



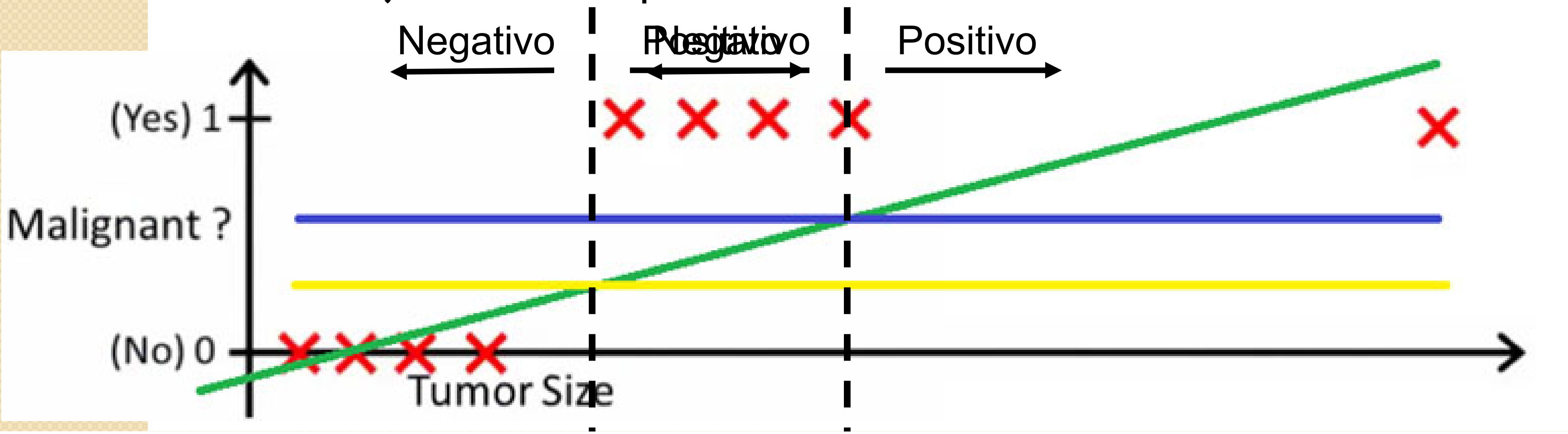
REGRESIÓN LOGÍSTICA

- Por qué no usar únicamente regresión lineal como modelo en problemas de clasificación
  - Ejemplo: Clasificación de tumores en malignos (positivo) o benignos (negativo) según el tamaño
    - Umbral en 0.5:



REGRESIÓN LOGÍSTICA

- Por qué no usar únicamente regresión lineal como modelo en problemas de clasificación
  - Ejemplo: Clasificación de tumores en malignos (positivo) o benignos (negativo) según el tamaño
    - Si el conjunto de datos tiene un valor atípico
      - La línea de mejor ajuste en la regresión lineal se desplaza para ajustarse a ese punto



Regresión Logística

- Por qué no usar únicamente regresión lineal como modelo en problemas de clasificación
  - Ejemplo: Clasificación de tumores en malignos (positivo) o benignos (negativo) según el tamaño
    - Si el conjunto de datos tiene un valor atípico
      - La línea de mejor ajuste en la regresión lineal se desplaza para ajustarse a ese punto
      - La línea azul: umbral antiguo
      - Línea amarilla: nuevo umbral, en este caso podría ser de 0.2
        - Sólo bajando el umbral dará resultados correctos
        - Por lo tanto, la regresión lineal es sensible a los valores atípicos

REGRESIÓN LOGÍSTICA

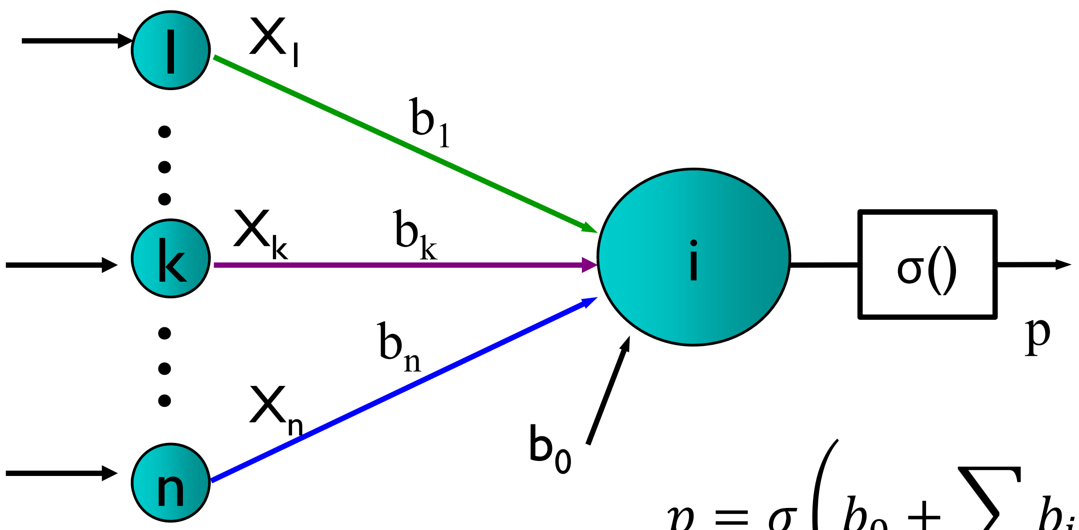
- Por qué no usar únicamente regresión lineal como modelo en problemas de clasificación
  - Además, un modelo de regresión lineal devuelve valores que no están acotados entre 0 y 1

REGRESIÓN LOGÍSTICA

- Perceptrón vs. Regresión logística
  - Regresión logística:

$$p = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-logit(p)}} = \frac{1}{1+e^{-(b_0+\sum b_i x_i)}}$$

  - Perceptrón con función de activación sigmoide:



$$p = \sigma\left(b_0 + \sum b_i x_i\right) = \frac{1}{1+e^{-(b_0+\sum b_i x_i)}}$$

Regresión Logística

- Entrenamiento
  - El objetivo del entrenamiento es encontrar los valores óptimos de los coeficientes b_0, b_1, b_2, … que minimicen el error de predicción
    - Es decir, que maximicen la probabilidad de que los datos observados sean generados por el modelo
  - Se utiliza la técnica de máxima verosimilitud para ajustar los coeficientes

REGRESIÓN LOGÍSTICA

- Entrenamiento
  - En teoría de probabilidad y estadística, la distribución Bernoulli (o distribución dicotómica) es una distribución de probabilidad discreta, dónde el valor 1 (éxito) ocurre con la probabilidad p y el valor 0 (fracaso) con la probabilidad 1-p
  - La función de probabilidad es

$$p(Y=y)=p^{y}(1-p)^{1-y} \qquad y\in\{0,1\}$$

    - Es decir

$$p(Y=y)=\begin{cases} p & y=1 \\ 1-p & y=0 \end{cases}$$

Regresión Logística

- Entrenamiento
  - Por lo tanto, con la función sigmoide, para una observación y:

$$p(Y = y|X = x) = p^y\ (1 - p)^{1-y}$$

  - Para n observaciones:
    - *Likelihood*

$$L = \prod p^y\ (1 - p)^{1-y}$$

    - Se desea maximizar esta expresión

con

$$p = \sigma(z) = \frac{1}{1 - e^{-z}} = \frac{1}{1 - e^{-(b_0 + \sum b_i x_i)}}$$

Regresión Logística

- Entrenamiento
  - Se toma el logaritmo en ambos lados:

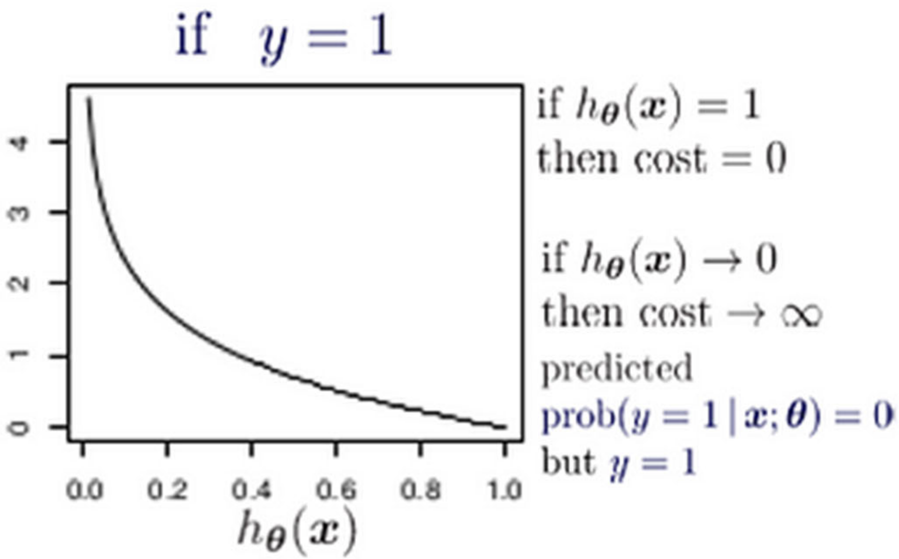
$$\mathrm{LL} = \log(L) = \sum [y \cdot \log(p) + (1-y) \cdot log(1-p)]$$

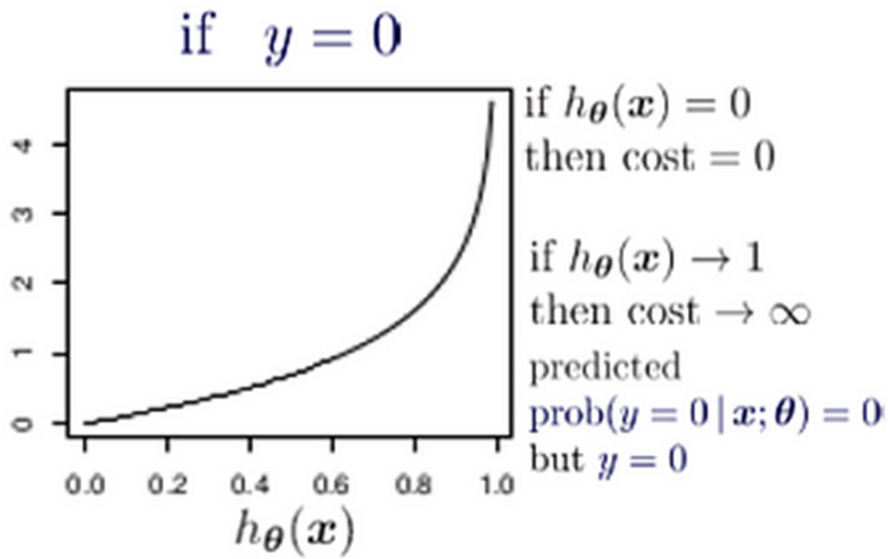
    - *Log-likelihood*
    - Se maximiza log(L) → se maximiza L
  - Para maximizar esta función, se aplicaría el ascenso de gradiente
    - En su lugar, se aplica el descenso de gradiente del negativo de esa función
      - Es decir, max(log(x)) = min(-log(x))
      - Maximización se transforma en minimización
      - El negativo de esta función es la función de coste

REGRESIÓN LOGÍSTICA

- Entrenamiento
  - Por tanto, la función de pérdida es:

$$\operatorname{cost}\left(h_{\boldsymbol{\theta}}(\boldsymbol{x}),\ y\right)=\begin{cases}-\log(h_{\boldsymbol{\theta}}(\boldsymbol{x})) & \text{if } \ y=1\\ -\log(1-h_{\boldsymbol{\theta}}(\boldsymbol{x})) & \text{if } \ y=0\end{cases}$$

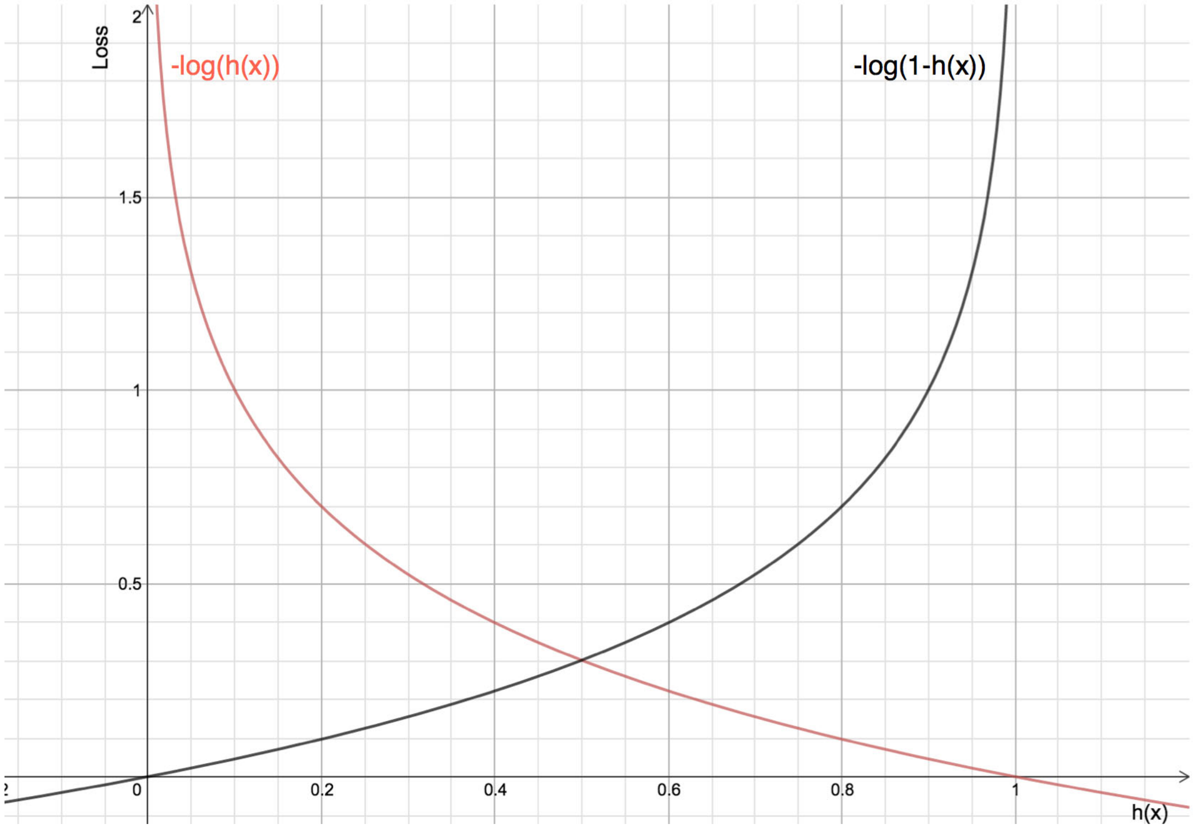




- $h_\theta(x)=\sigma(x)$: predicción
  - Θ: conjunto de coeficientes (b_i)

REGRESIÓN LOGÍSTICA

- Entrenamiento
  - Si se combinan ambos gráficos, se obtiene una función con un único mínimo:



Regresión Logística

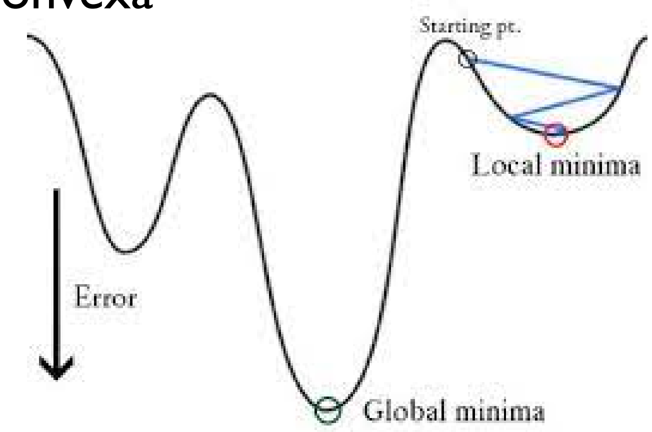
- Entrenamiento
  - Si se combinan ambos gráficos, se obtiene una función con un único mínimo:
    - Línea roja: clase positiva (y=1)
      - Si la probabilidad predicha se aproxima a 1, la pérdida será menor
      - Si la probabilidad se acerca a 0, la función de pérdida alcanzará el infinito
    - Línea negra: clase negativa 0 (y=0)
      - Si la probabilidad predicha es cercana a 0, la pérdida será menor
      - Si la probabilidad se aproxima a 1 entonces la función de pérdida alcanza el infinito

Regresión Logística

- Entrenamiento
  - Esta función de coste también se denomina *log loss*
    - Pérdida logarítmica
  - Garantiza que, a medida que se maximiza la probabilidad de la respuesta correcta, se minimiza la probabilidad de la respuesta incorrecta
    - Por tanto, cuanto menor sea el valor de esta función de coste, mayor será la precisión

Regresión Logística

- Entrenamiento
  - ¿Por qué no se usa la función de coste de regresión lineal?
    - Regresión lineal: usa error cuadrático medio
      - Se deriva del estimador de máxima verosimilitud
      - Tiene un único mínimo
    - Si se usa, habrá mínimos locales
      - Función no convexa



REGRESIÓN LOGÍSTICA

- Entrenamiento
  - Por tanto, se utiliza una función de coste diferente para regresión logística
    - Llamada *log loss*
    - También se deriva del método de estimación de máxima verosimilitud

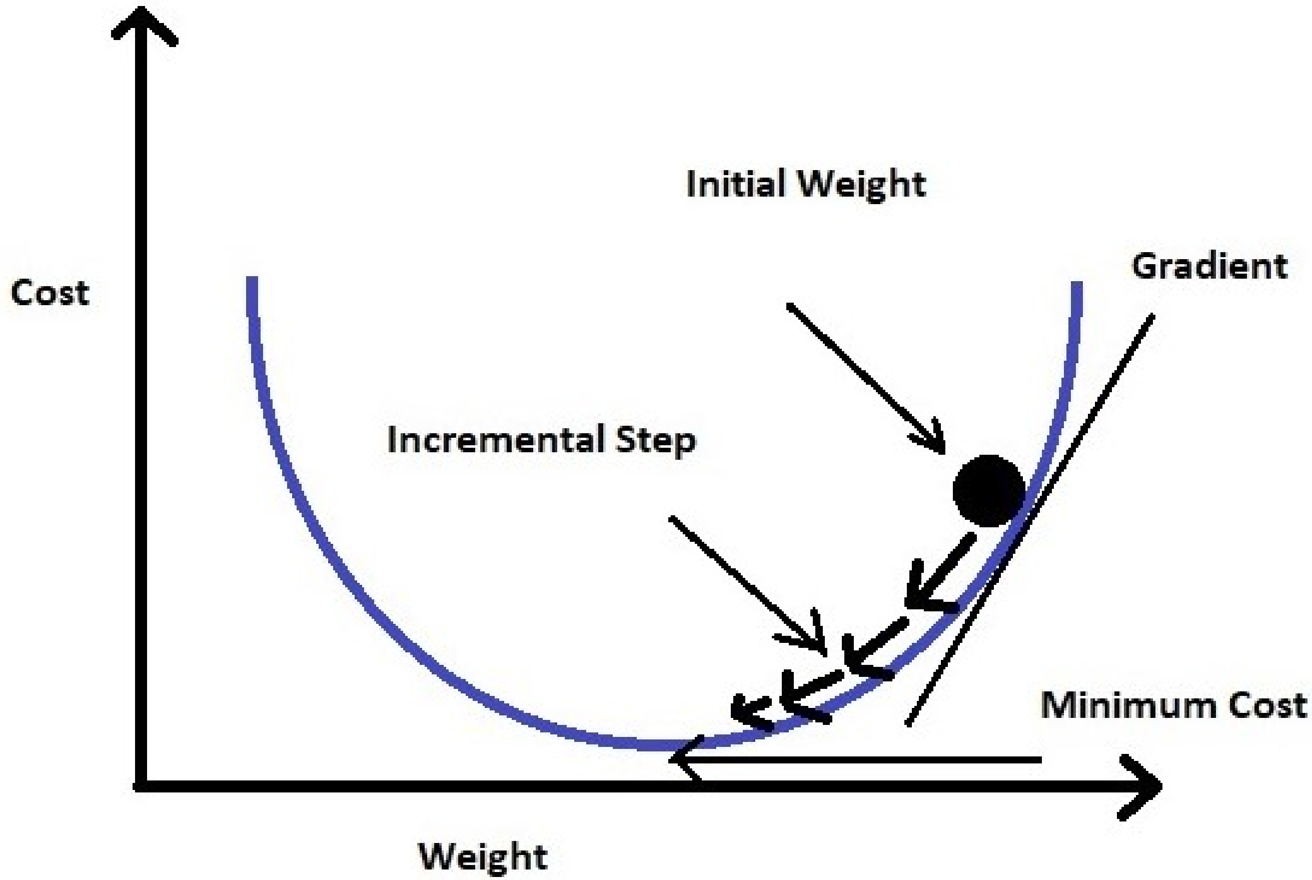
$$logloss = \frac{1}{N}\sum_{i=1}^{N} -[y_i \cdot log(p_i) + (1 - y_i) \cdot log(1 - p_i)]$$

Regresión Logística

- Entrenamiento
  - El descenso de gradiente cambia el valor de los coeficientes de tal manera que siempre converge al punto mínimo
    - Es decir, el objetivo es encontrar los coeficientes óptimos que minimizan la función de pérdida del modelo
  - Es un método iterativo que encuentra el mínimo de una función calculando la pendiente en un punto y luego moviéndose **en la dirección opuesta**

Regresión Logística

- Entrenamiento
  - Descenso de gradiente:



Regresión Logística

- Entrenamiento
  - Descenso de gradiente:

Minimizar el negativo de LL

$$b_i(t+1) = b_i(t) - \alpha \frac{\partial(-LL)}{\partial b_i} = b_i(t) + \alpha \frac{\partial LL}{\partial b_i}$$

Necesario moverse en el sentido contrario del gradiente

- Derivada: regla de la cadena:

$$\frac{\partial LL}{\partial b_i} = \frac{\partial LL}{\partial p}\frac{\partial p}{\partial b_i} = \frac{\partial LL}{\partial p}\frac{\partial p}{\partial z}\frac{\partial z}{\partial b_i}$$

con

$$p = \sigma(z) \qquad z = b_0 + \sum b_i x_i$$

Regresión Logística

- Entrenamiento
  - Cálculo de cada una de esas derivadas parciales:

$$\frac{\partial LL}{\partial b_i} = \boxed{\frac{\partial LL}{\partial p}} \frac{\partial p}{\partial z} \frac{\partial z}{\partial b_i}$$

$$\mathrm{LL} = \sum [y \cdot \log(p) + (1-y) \cdot log(1-p)] \Rightarrow \frac{\partial LL}{\partial p} = \frac{y}{p} - \frac{1-y}{1-p}$$

REGRESIÓN LOGÍSTICA

- Entrenamiento
  - Cálculo de cada una de esas derivadas parciales:

$$\frac{\partial LL}{\partial b_i} = \frac{\partial LL}{\partial p} \boxed{\frac{\partial p}{\partial z}} \frac{\partial z}{\partial b_i}$$

$p = \sigma(z)$

$$\frac{dp}{dz} = \frac{d\sigma}{dz} = \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right) = \frac{-e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \frac{e^{-z}}{1+e^{-z}}$$

$$= \frac{1}{1+e^{-z}}\left(1 - \frac{e^{-z}}{1+e^{-z}}\right) = \sigma(z)\big(1-\sigma(z)\big) = p(1-p)$$

REGRESIÓN LOGÍSTICA

- Entrenamiento
  - Cálculo de cada una de esas derivadas parciales:

$$\frac{\partial LL}{\partial b_i} = \frac{\partial LL}{\partial p}\frac{\partial p}{\partial z}\boxed{\frac{\partial z}{\partial b_i}}$$

$$z = b_0 + \sum b_i x_i \quad \Rightarrow \quad \frac{\partial z}{\partial b_i} = \frac{\partial}{\partial b_i}\left(b_0 + \sum b_i x_i\right) = x_i$$

Regresión Logística

- Entrenamiento
  - Uniendo las tres derivadas parciales:

$$\frac{\partial LL}{\partial b_i} = \frac{\partial LL}{\partial p}\frac{\partial p}{\partial z}\frac{\partial z}{\partial b_i} =$$

$$\frac{\partial LL}{\partial p} = \frac{y}{p} - \frac{1-y}{1-p} \qquad \frac{dp}{dz} = p(1-p) \qquad \frac{\partial z}{\partial b_i} = x_i$$

$$= \left[\frac{y}{p} - \frac{1-y}{1-p}\right](p(1-p))x_i =$$

$$= (y(1-p) - (1-y)p)x_i = (y-p)x_i$$

REGRESIÓN LOGÍSTICA

- Entrenamiento
  - Por lo tanto, la regla de modificación de b_i es:

$$\left.\begin{aligned} b_i(t+1) &= b_i(t) + \alpha \frac{\partial LL}{\partial b_i} \\ \frac{\partial LL}{\partial b_i} &= (y-p)x_i \end{aligned}\right\} b_i(t+1) = b_i(t) + \alpha(y-p)x_i$$

Regresión Logística

- Entrenamiento
  - Comparando esta expresión con la Regla Delta:

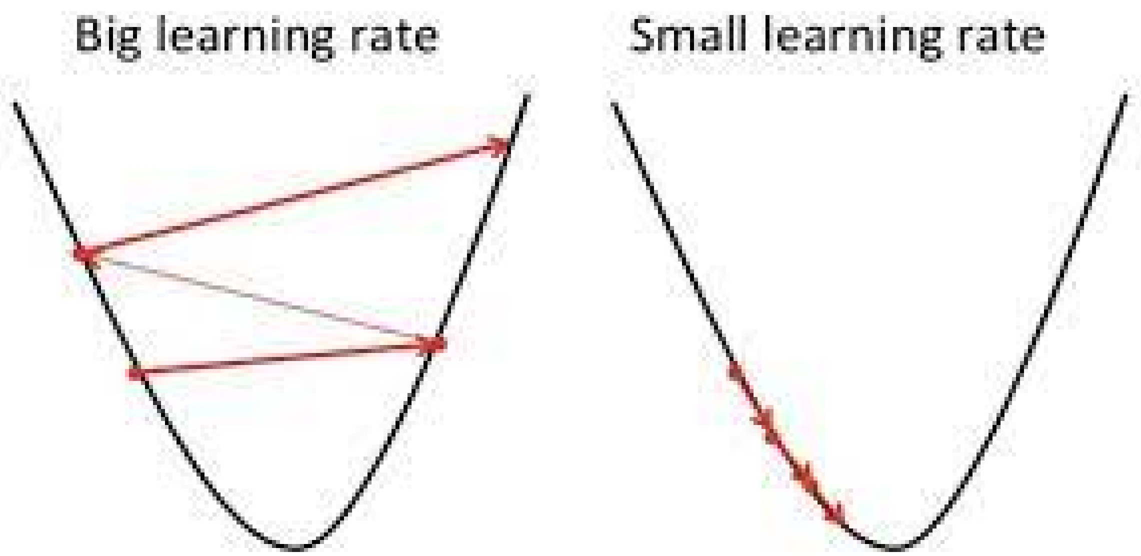
$$b_i(t+1) = b_i(t) + \alpha(y-p)x_i$$

$$w_j(t+1) = w_j(t) + \mu(d-y)x_j$$

    - Expresiones idénticas
      - Partiendo de funciones de coste distintas

REGRESIÓN LOGÍSTICA

- Entrenamiento
  - Descenso de gradiente:
    - α: tasa de aprendizaje
      - Mismo comportamiento que en la Regla Delta



Regresión Logística

- Función de pérdida vs. Función de coste
  - *Loss function* vs. *Cost function*
  - Muchos autores usan ambos términos como sinónimos
  - Muchas técnicas de AA se basan en una función de pérdida/coste que minimizar
    - Por ejemplo: Funciones de coste de RR.NN.AA.
      - Regresión: Error cuadrático medio
      - Clasificación: entropía cruzada

Regresión Logística

- Función de pérdida vs. Función de coste
  - La función de pérdida (o error) es para un ejemplo
    - Es un vector con un valor por cada ejemplo
  - La función de coste es una medida con respecto a todo el conjunto de entrenamiento
    - O el lote (*batch*) utilizado para entrenar
    - Es un valor único, no un vector
      - Evalúa lo bien que lo ha hecho el modelo en su conjunto.

REGRESIÓN LOGÍSTICA

- Función de pérdida vs. Función de coste
  - Por tanto, la formulación de la función de coste se basa en la función de pérdida
    - Comúnmente, la suma o el promedio de la función de pérdida
  - Además, la función de coste puede contener términos de **regularización** además de la función de pérdida
    - Para evitar el sobreajuste

REGRESIÓN LOGÍSTICA

- Regularización
  - Evitar el sobreajuste
  - Técnicas comunes en Regresión Logística:
    - Regularización L1 (*Lasso*)
    - Regularización L2 (*Ridge*)

    - Provenientes de la regresión lineal
    - No son propias de regresión logística
      - Es común utilizarlas en otras técnicas
        - Por ejemplo, muy utilizadas en RR.NN.AA.

Regresión Logística

- Regularización
  - Regularización L1 (*Lasso*) y L2 (*Ridge*)
    - Idea principal:
      - A lo largo del proceso de entrenamiento es habitual que ciertos parámetros tomen valores muy altos para ajustarse a ejemplos concretos del conjunto de entrenamiento
        - Es decir, particulariza en esos ejemplos
      - La presencia de valores grandes suele provocar que el modelo produzca grandes variaciones en la salida para pequeñas variaciones en la entrada
        - Como resultado, los valores grandes suelen tener un efecto adverso en la capacidad de generalización de una red
        - Los valores altos en parámetros suelen ser característicos de un modelo sobreajustado
      - Mantener los valores bajos añadiendo una penalización al *loss*

Regresión Logística

- Regularización
  - Regularización L1 (*Lasso*)
    - Lasso es la abreviatura de *Least Absolute Shrinkage and Selection Operator*
    - Se utiliza tanto para la regularización como para la selección de características
    - Esta técnica agrega a la función de coste un término de penalización que es proporcional al valor absoluto de los coeficientes (norma L_1)

REGRESIÓN LOGÍSTICA

- Regularización
  - Regularización L1 (*Lasso*)
    - Función de coste para regresión lineal:

$$L = \sum (y_i - \hat{y}_i)^2 = \sum \underbrace{\left(y_i - \sum b_j x_j\right)^2}_{\text{Función de pérdida}} \Rightarrow L = \sum \underbrace{\left(y_i - \sum b_j x_j\right)^2}_{\text{Función de pérdida}} + \underbrace{\lambda \sum |b_j|}_{\text{Regularización}}$$

REGRESIÓN LOGÍSTICA

- Regularización
  - Regularización L1 (*Lasso*)
    - Función de coste para regresión lineal:

$$L = \sum (y_i - \hat{y}_i)^2 = \sum \left(y_i - \sum b_j x_j \right)^2 \quad \Rightarrow \quad L = \sum \left(y_i - \sum b_j x_j \right)^2 + \lambda \sum |b_j|$$

    - λ es el término de penalización que denota la cantidad de contracción (o restricción) que se implementará en la ecuación.
      - λ puede ser cualquier número de valor real entre cero e infinito
        - Cuanto mayor sea el valor, más agresiva será la penalización
          - Un valor mayor penaliza la función de optimización
        - Con λ=0, se tiene el equivalente del modelo de regresión lineal de la ecuación

REGRESIÓN LOGÍSTICA

- Regularización
  - Regularización L1 (*Lasso*)
    - Debido a que los coeficientes se "encogen", algunos tomarán un valor de 0
      - L1 tiende a forzar que algunos coeficientes se vuelvan igual a 0
        - Esto equivale a eliminar las características correspondientes
          - Es decir, esta técnica permite realizar **selección automática de características**
        - Además, esto conlleva una simplificación del modelo
          - Más sencillo → menos sobreajuste → mayor generalización
    - Es decir, la regularización *lasso* reduce los coeficientes y ayuda a reducir la complejidad del modelo

REGRESIÓN LOGÍSTICA

- Regularización
  - Regularización L2 (*Ridge*)
    - Similar a L1, se añade una restricción a los coeficientes mediante la introducción de un factor de penalización
      - En este caso, la penalización es la suma del cuadrado de los coeficientes
    - Función de coste para regresión lineal:

$$L = \sum (y_i - \hat{y}_i)^2 = \sum \left(y_i - \sum b_j x_j \right)^2 \Rightarrow L = \sum \left(y_i - \sum b_j x_j \right)^2 + \lambda \sum b_j^2$$

    - Ridge reduce la magnitud de los coeficientes sin necesariamente llevarlos a cero

Regresión Logística

- Regularización
  - Regularización L2 (*Ridge*)
    - A diferencia de L1, *Ridge* no tiende a anular coeficientes
      - **No permite realizar selección automática de características**
    - En lugar de ello reduce su magnitud de manera uniforme
    - Ayuda a evitar que los coeficientes se vuelvan demasiado grandes y **aporta estabilidad al modelo**
    - En redes de neuronas, este tipo de regularización también se llama decaimiento de pesos (*weight decay*)

Regresión Logística

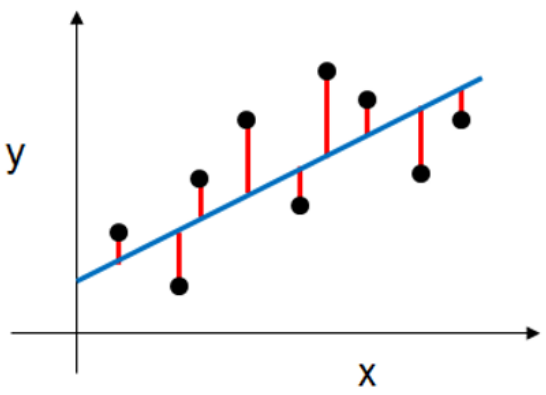
- Regularización
  - Regularización L1 (*Lasso*) y L2 (*Ridge*)
    - Si un modelo de regresión lineal utiliza la técnica de regularización L1, se denomina **regresión *lasso***
    - Si un modelo de regresión lineal utiliza la técnica de regularización L2, se denomina **regresión *ridge***

Regresión Logística

- Regularización
  - Regularización L1 (*Lasso*) y L2 (*Ridge*)

Ridge regression
(or "L2 regularization") minimizes:

$$\text{RSS} + \alpha \sum_{j=1}^{p} \beta_j^2$$



Model Prediction

$$SS_{residuals} = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

Observed Result

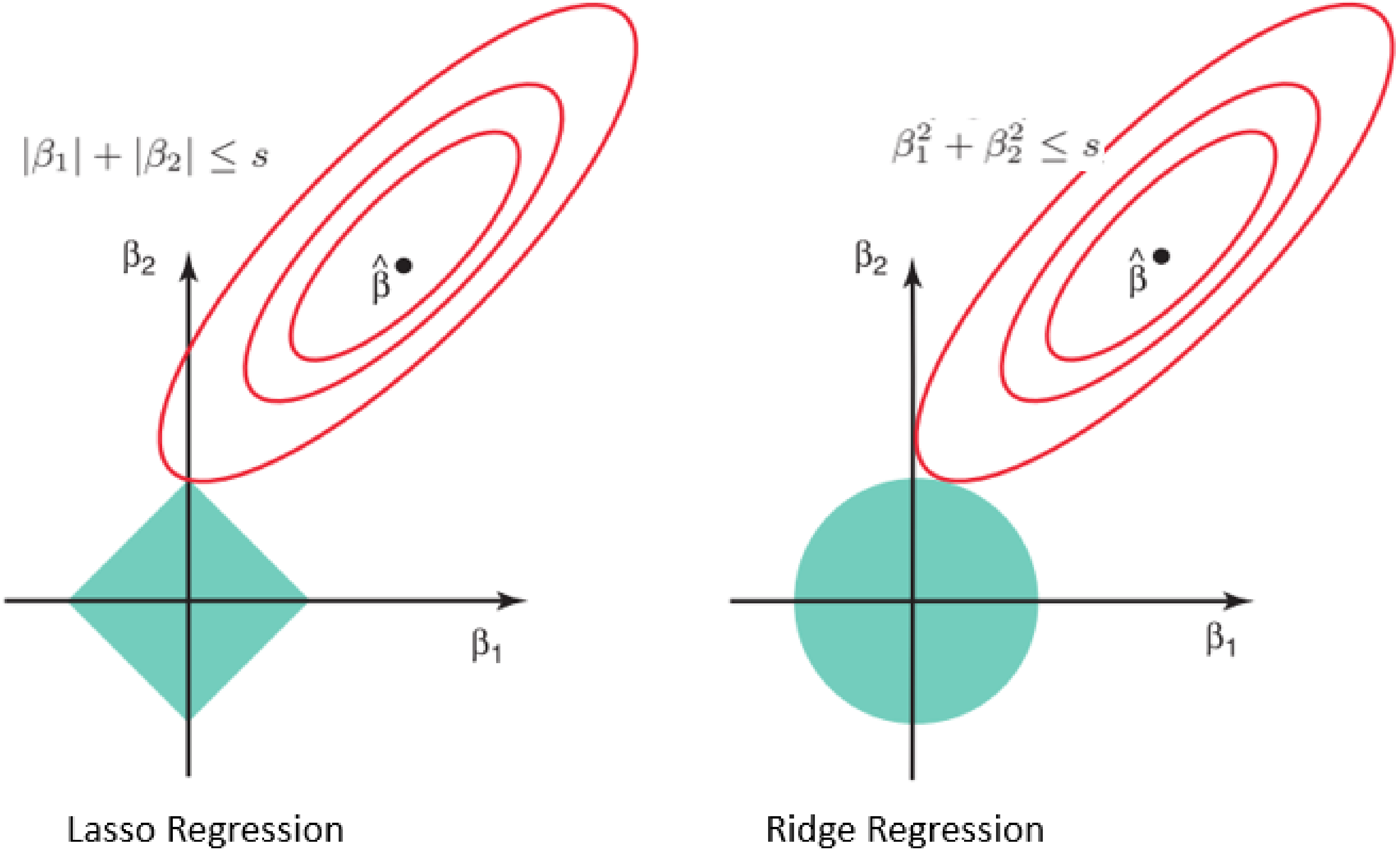
$$\text{RSS} + \alpha \sum_{j=1}^{p} |\beta_j|$$

Lasso regression
(or "L1 regularization") minimizes:

p is the number of features
βj is a model coefficient
α is a tuning parameter: Increasing the α penalizes the coefficients and thus shrinks them.

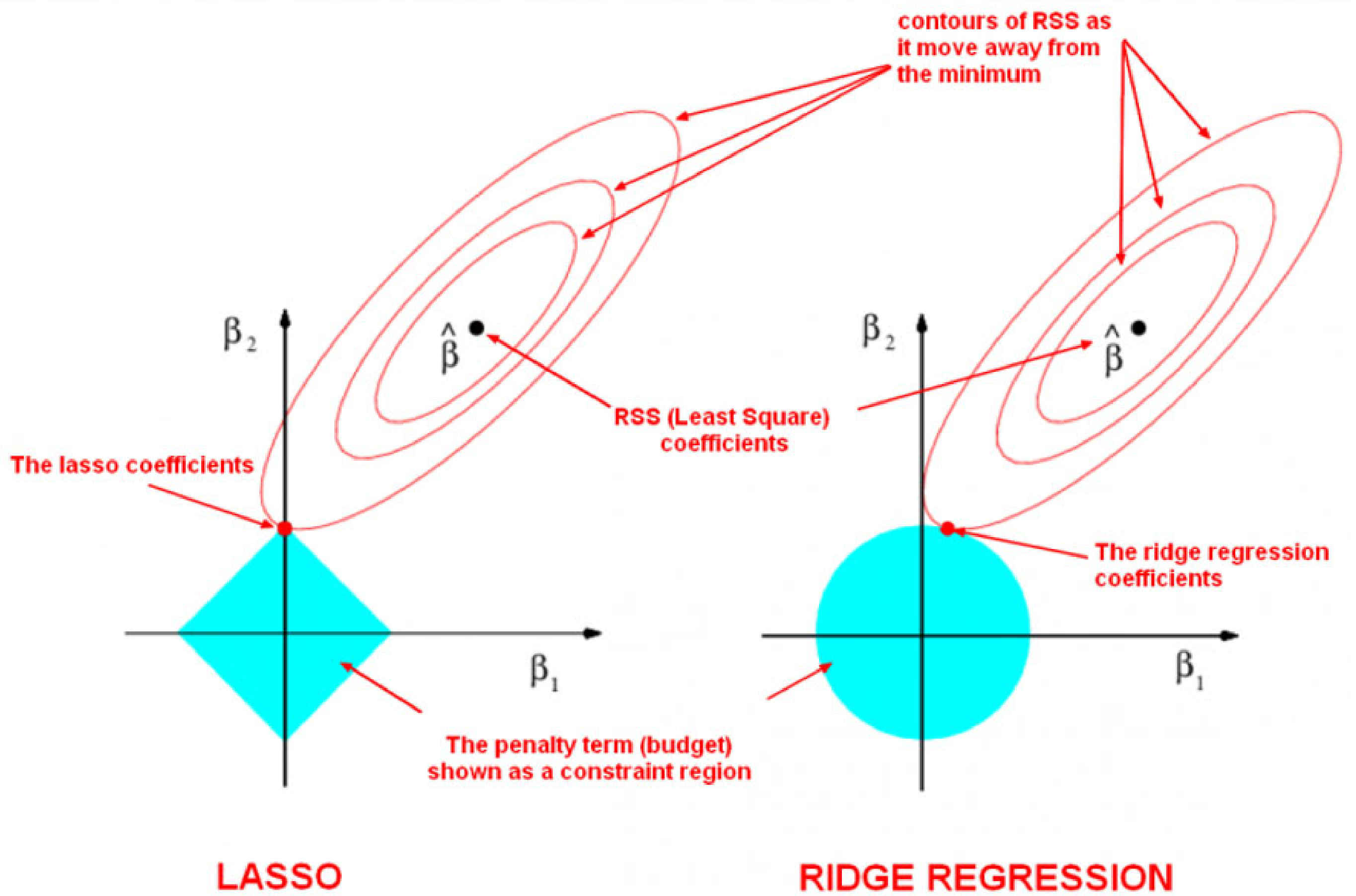
REGRESIÓN LOGÍSTICA

- Regularización
  - Regularización L1 (*Lasso*) y L2 (*Ridge*)



REGRESIÓN LOGÍSTICA

- Regularización
  - Regularización L1 (*Lasso*) y L2 (*Ridge*)



REGRESIÓN LOGÍSTICA

- Regularización
  - Regularización L1 (*Lasso*) y L2 (*Ridge*)
    - Contornos elípticos (rojos): funciones de coste
      - Una elipse para un valor de λ
        - λ cercano a 0: elipse menor
          - λ=0: punto equivalente a regresión lineal
        - λ alto: elipse mayor
    - Rombo, círculo: región restringida de cada modelo
      - La penalización hace que la elipse no pueda estar en el interior
    - Ambos métodos determinan los coeficientes encontrando el primer punto en el que los contornos elípticos chocan con la región de restricciones

Regresión Logística

- Regularización
  - Regularización L1 (*Lasso*) y L2 (*Ridge*)
    - La relajación de las restricciones introducidas por el factor de penalización conduce a un aumento de la región restringida
      - Rombo en *Lasso*, círculo en *Ridge*
      - Es decir, disminuyendo λ
      - Si se disminuyendo λ, se llegará al centro de la elipse
        - En este punto, los resultados tanto de los modelos lasso como ridge son similares a los de un modelo de regresión lineal

Regresión Logística

- Regularización
  - Regularización L1 (*Lasso*) y L2 (*Ridge*)
    - Regresión *lasso*: la región con restricciones tiene una forma de diamante
      - Por tanto, cada vez que interseca con una región elíptica, al menos uno de los coeficientes se convierte en cero
        - Es decir, al menos una de las características se anulará
          - Se multiplicará por 0
    - Regresión *ridge*: forma circular
      - Por tanto, esto no ocurre
        - Es decir, los valores pueden reducirse cerca de cero, pero nunca ser iguales a cero
          - Por lo tanto, **no se puede utilizar para hacer una selección de características**

Regresión Logística

- Regularización
  - Otras técnicas de regularización:
    - Elastic Net
    - Regularización de Tikhonov (RidgeCV y LassoCV)
    - Regularización por Norma Max (Max-Norm Regularization)
    - Dropout como Regularización
    - Parada temprana (*Early Stopping*)

Regresión Logística

- Regularización
  - Elastic Net
    - Combina las regularizaciones L1 y L2, incorporando ambas penalizaciones en la función de coste
    - Esto proporciona un equilibrio entre
      - Selección de características (como en *Lasso*)
      - Estabilidad de los coeficientes (como en *Ridge*)

    - En general, Elastic Net no es una técnica de regularización, sino un modelo de regresión lineal con una función de coste que incorpora penalizaciones L1 y L2

REGRESIÓN LOGÍSTICA

- Regularización
  - Elastic Net
    - Es decir, Elastic Net es una combinación de dos de los mejores enfoques de regresión lineal por "contracción":
      - Regresión *lasso* (penalización L1)
        - Se ocupa de la selección de características de los coeficientes de regresión
        - Para ello, utiliza la contracción de los coeficientes de regresión hacia cero o igual a cero para reducir la aparición de variables predictoras
      - Regresión *ridge* (penalización L2)
        - Se ocupa de los problemas de alta multicolinealidad

Regresión Logística

- Regularización
  - Elastic Net
    - Función de coste utilizada:

$$L = \sum \left(y_i - \sum b_j x_j \right)^2 + \lambda_1 \sum |b_j| + \lambda_2 \sum b_j^2$$

    - Más comúnmente, se suele utilizar (equivalente):

$$L = \sum \left(y_i - \sum b_j x_j \right)^2 + \alpha \left(\lambda \sum |b_j| + (1 - \lambda) \sum b_j^2 \right)$$

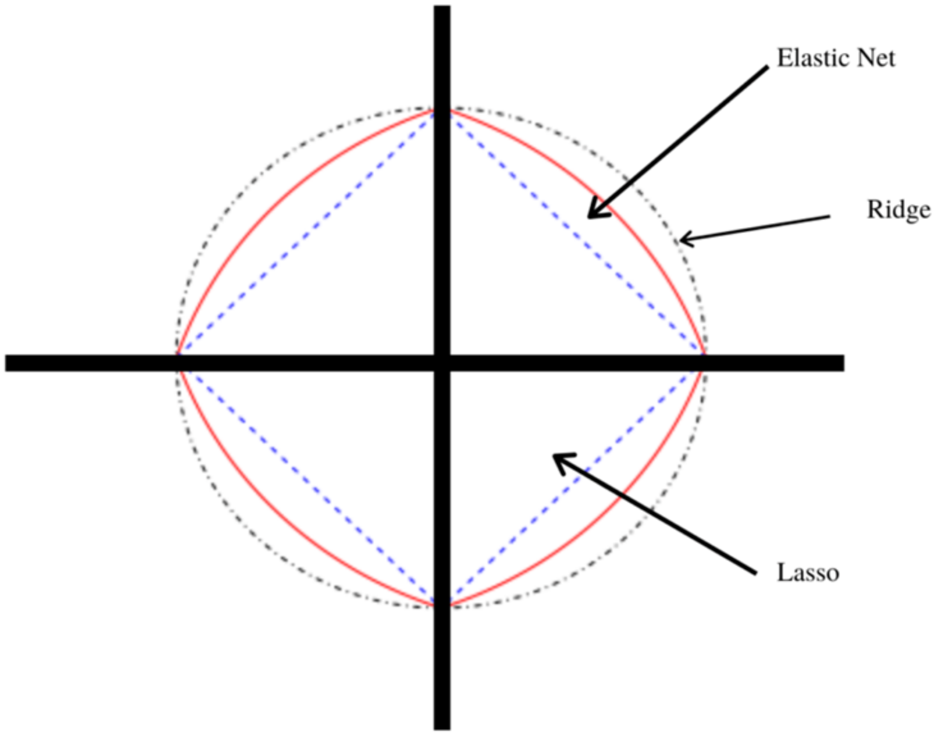
      - donde λ es el ratio de L1 y α es el nivel de penalización global

REGRESIÓN LOGÍSTICA

- Regularización
  - Elastic Net
    - Ayuda a producir un modelo ajustado e interpretable minimizando las variables innecesarias que no aparecen en el modelo final para mejorar la precisión de la predicción
    - Gestiona la multicolinealidad manteniendo o excluyendo del modelo ajustado las variables predictoras altamente correlacionadas

REGRESIÓN LOGÍSTICA

- Regularización
  - Elastic Net
    - Elastic Net (rojo) es una combinación de la regresión *ridge* (negro) y *lasso* (azul)



REGRESIÓN LOGÍSTICA

- Regularización
  - Elastic Net
    - Regresión lineal, *lasso* y *ridge* se pueden considerar casos especiales de Elastic Net
    - En 2014, se demostró que Elastic Net puede reducirse a una SVM lineal
    - La función de *loss* es fuertemente convexa y, por tanto, existe un mínimo único

Regresión Logística

- Regularización
  - Regularización de Tikhonov
    - Un problema de regresión lineal intenta encontrar un vector **x** tal que

$$A\mathbf{x} = \mathbf{b}$$

    - mediante la minimización de

$$\|A\mathbf{x} - \mathbf{b}\|_2^2$$

      - $\| \cdot \|_2$: norma euclídea
    - Sin embargo, para hallar una solución que tenga unas propiedades específicas, se añade un término de regularización a la función a minimizar:

$$\|A\mathbf{x} - \mathbf{b}\|_2^2 + \|\Gamma\mathbf{x}\|_2^2$$

REGRESIÓN LOGÍSTICA

- Regularización
  - Regularización de Tikhonov
    - Si se escoge una matriz de Tikhonov Γ que sea un múltiplo escalar de la matriz identidad ($\Gamma = \alpha I$)
      - Se intentará encontrar soluciones con normas bajas
      - Esto es la regresión *ridge*

    - Es decir, la regularización L2 (*ridge*) es un caso particular de esta

Regresión Logística

- Regularización
  - Regularización por Norma Max (*Max-Norm Regularization*)
    - Idea principal:
      - Similar a las técnicas anteriores
        - Evitar que los parámetros tomen valores muy altos
          - Para que el modelo no se ajuste a ejemplos específicos del conjunto de entrenamiento
      - Técnica de regularización algo más agresiva que L1 y L2 a la hora de evitar el uso de pesos grandes
        - L1 y L2 → Añaden un término de penalización
          - «Desincentivan» el uso de valores grandes
        - *Max-Norm* → Restringe los valores, garantizando que su magnitud no supere un determinado valor umbral

Regresión Logística

- Regularización
  - Regularización por Norma Max (*Max-Norm Regularization*)

    - Por tanto, esta técnica se implementa imponiendo un límite superior a la magnitud del vector de parámetros
      - Esta magnitud suele ser medida con la norma L2
        - Pero podría ser con cualquier otra norma
      - Si la norma L2 del vector de coeficientes supera este límite, se escala el vector para que cumpla con el límite

REGRESIÓN LOGÍSTICA

- Regularización
  - Regularización por Norma Max (*Max-Norm Regularization*)
    - Esta técnica es particularmente útil en redes neuronales profundas
      - El control de la magnitud de los parámetros puede contribuir a un entrenamiento más estable y prevenir problemas de explosión del gradiente
      - También puede ayudar a mejorar la capacidad de generalización del modelo al limitar su complejidad

$$w = \begin{cases} w & \text{if } ||w||_2 \leq r \\ r \frac{w}{||w||_2} & \text{otherwise.} \end{cases}$$

$$||w||_2 = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}$$

Regresión Logística

- Regularización
  - *Dropout* como Regularización
    - Aunque generalmente se asocia con redes neuronales, el *dropout* también puede aplicarse a regresión logística y otros modelos
      - *Dropout* en RR.NN.AA.:
        - Desactivar aleatoriamente varias neuronas en cada ciclo (*epoch*) de entrenamiento

Regresión Logística

- Regularización
  - *Dropout* como Regularización
    - Aunque generalmente se asocia con redes neuronales, el *dropout* también puede aplicarse a regresión logística y otros modelos
      - *Dropout* en otros modelos:
        - Desactivar aleatoriamente un porcentaje de características (coeficientes) durante cada iteración de entrenamiento
        - Sólo es válido en modelos que se entrenan de forma iterativa, modificando los parámetros de un único modelo en cada iteración
          - Por ejemplo, en Computación Evolutiva no se podría utilizar
            - Es iterativo, pero cada individuo es un modelo en sí
              - No se modifica el mismo modelo en cada iteración

Regresión Logística

- Regularización
  - Parada temprana (*early stopping*)
    - Muy utilizado en RR.NN.AA., pero también puede ser utilizado en otros modelos
    - Detener el entrenamiento del modelo cuando el rendimiento en un conjunto de validación comienza a empeorar
      - Porque se considera que empieza a empeorar en funcionamiento con datos nuevos (validación)
        - Empieza a sobreajustar
    - Además, el modelo a devolver puede no ser el que se tiene en el momento de parar el entrenamiento
      - Puede ser el de un ciclo anterior: el de menor validación

REGRESIÓN LOGÍSTICA

- Regularización
  - Parada temprana (*early stopping*)
    - Esto puede ayudar a evitar que el modelo se ajuste demasiado a los datos de entrenamiento
      - Se intenta que el modelo «no desarrolle» su complejidad
    - Al igual que antes, sólo puede ser aplicado en modelos cuyo entrenamiento es iterativo
      - Para poder parar antes en una iteración temprana
      - Si se aplica en modelos entrenados en un proceso de evaluación de muchos modelos
        - Como en Computación Evolutiva
        - ¡Se podría sobreajustar al conjunto de validación!

REGRESIÓN LOGÍSTICA

- Ventajas:
  - Intrínsecamente interpretable
    - Se puede entender cómo las características afectan las probabilidades de clasificación
    - Es decir, los coeficientes $b_1,b_2,\ldots,b_n$ en la ecuación del modelo tienen significado
      - Un coeficiente positivo indica que un aumento en la característica incrementa la probabilidad de pertenecer a la clase 1
      - Un coeficiente negativo indica lo contrario.
  - Entrenamiento eficiente
    - Basado en gradiente descendente

Regresión Logística

- Desventajas:
  - El *log-odds* se basa en un modelo lineal
    - Capacidad de separar datos muy limitada
      - Hiperplano
      - Muchos problemas del mundo real requieren de regiones de separación más complejas

REGRESIÓN LOGÍSTICA

- Conclusiones
  - Técnica de clasificación que combina la regresión lineal con una función sigmoide para calcular la probabilidad de pertenencia a la clase positiva
    - Regresión lineal → realiza una separación lineal
  - Se puede entrenar mediante gradiente descendente
    - Necesario definir su propia función de coste
    - Expresión final idéntica a la Regla Delta
  - Aplicación de técnicas de regularización
    - Nuevos modelos de regresión
      - Regresión *lasso*, *ridge*, *elastic net*