Scuola universitaria professionale della Svizzera italiana
Bachelor di Ingegneria Informatica

SUPSI

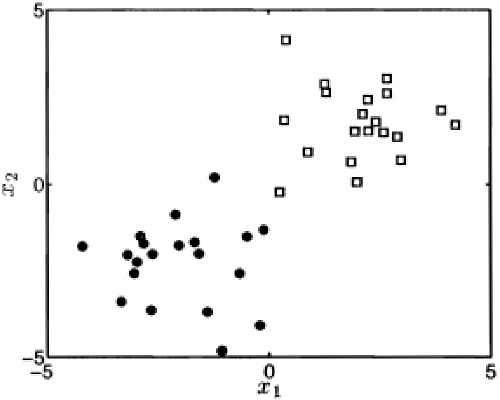
Machine Learning

Lezione 5 - Regressione Logistica

Loris Cannelli, Ricercatore, IDSIA-SUPSI
loris.cannelli@supsi.ch

IDSIA-SUPSI, Polo universitario Lugano - Dipartimento Tecnologie Innovative

Classificazione Binaria



- Due attributi x_1 e x_2
- Target binario $t = \{0, 1\}$

$\Rightarrow$ Classificazione

Classificazione Binaria

Assumiamo che la variabile osservata sia una variabile aleatoria di tipo *Bernoulli*

Classificazione Binaria

Assumiamo che la variabile osservata sia una variabile aleatoria di tipo *Bernoulli*

- $t \sim \text{Bernoulli}(p)$ può avere come esiti $\{0, 1\}$

Classificazione Binaria

Assumiamo che la variabile osservata sia una variabile aleatoria di tipo *Bernoulli*

- $t \sim \text{Bernoulli}(p)$ può avere come esiti $\{0, 1\}$
  - Esempio: Lancio di una moneta

Classificazione Binaria

Assumiamo che la variabile osservata sia una variabile aleatoria di tipo *Bernoulli*

- $t \sim \text{Bernoulli}(p)$ può avere come esiti $\{0, 1\}$
  - Esempio: Lancio di una moneta
- p *probabilità di successo*, cioè $p = P(t = 1)$

Classificazione Binaria

Assumiamo che la variabile osservata sia una variabile aleatoria di tipo *Bernoulli*

- $t \sim \text{Bernoulli}(p)$ può avere come esiti $\{0, 1\}$
  - Esempio: Lancio di una moneta
- p *probabilità di successo*, cioè $p = P(t = 1)$
  - Esempio: $p = \frac{1}{2} = P(\text{testa})$

Classificazione Binaria

Assumiamo che la variabile osservata sia una variabile aleatoria di tipo *Bernoulli*

- $t \sim \text{Bernoulli}(p)$ può avere come esiti $\{0, 1\}$
  - Esempio: Lancio di una moneta
- p *probabilità di successo*, cioè $p = P(t = 1)$
  - Esempio: $p = \frac{1}{2} = P(\text{testa})$
- Funzione di probabilità:

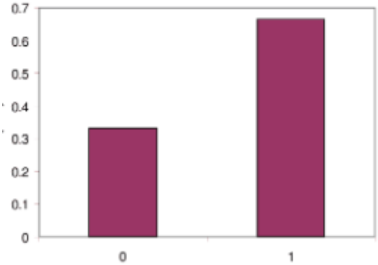
$$\text{Bernoulli}(t; p) = \begin{cases} 1 - p, & \text{for } t = 0 \\ p, & \text{for } t = 1 \end{cases}$$

Classificazione Binaria

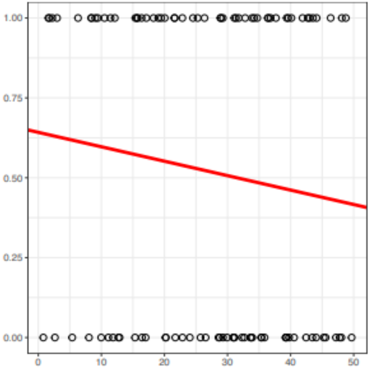
Assumiamo che la variabile osservata sia una variabile aleatoria di tipo *Bernoulli*

- $t \sim \text{Bernoulli}(p)$ può avere come esiti $\{0, 1\}$
  - Esempio: Lancio di una moneta
- p *probabilità di successo*, cioè $p = P(t = 1)$
  - Esempio: $p = \frac{1}{2} = P(\text{testa})$
- Funzione di probabilità:

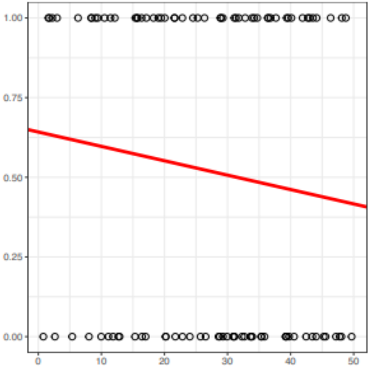
$$\text{Bernoulli}(t; p) = \begin{cases} 1 - p, & \text{for } t = 0 \\ p, & \text{for } t = 1 \end{cases}$$



Classificazione Binaria

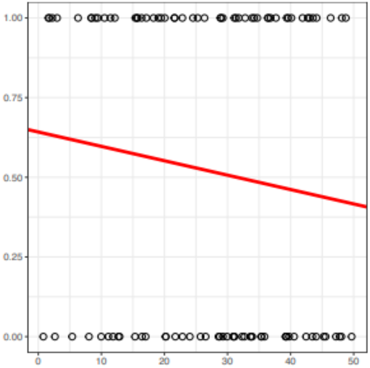


Classificazione Binaria



- La relazione tra x e t non è sicuramente lineare

Classificazione Binaria



- La relazione tra x e t non è sicuramente lineare
- La regressione lineare produce stime in $(-\infty, +\infty) \neq \{0, 1\}$

Sigmoide

$\Rightarrow$ Invece di stimare t, proviamo a stimare $P(t = 1|\mathbf{x})$, cioè la probabilità di ottenere un successo (cioè, $t = 1$), conoscendo i dati $\mathbf{x}$

Sigmoide

$\Rightarrow$ Invece di stimare t, proviamo a stimare $P(t = 1|\mathbf{x})$, cioè la probabilità di ottenere un successo (cioè, $t = 1$), conoscendo i dati $\mathbf{x}$

- Dato che vogliamo stimare una probabilità, ci serve una qualche funzione che rimanga tra 0 e 1

Sigmoide

$\Rightarrow$ Invece di stimare t, proviamo a stimare $P(t = 1|\mathbf{x})$, cioè la probabilità di ottenere un successo (cioè, $t = 1$), conoscendo i dati $\mathbf{x}$

- Dato che vogliamo stimare una probabilità, ci serve una qualche funzione che rimanga tra 0 e 1
- Introduciamo il sigmoide:

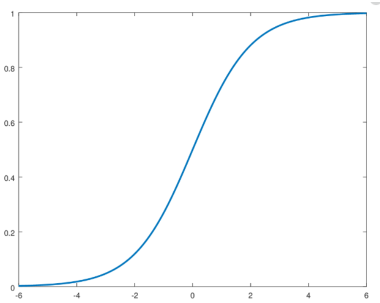
$$\sigma(a) \triangleq \frac{1}{1 + e^{-a}} \in (0, 1)$$

Sigmoide

$\Rightarrow$ Invece di stimare t, proviamo a stimare $P(t = 1|\mathbf{x})$, cioè la probabilità di ottenere un successo (cioè, $t = 1$), conoscendo i dati $\mathbf{x}$

- Dato che vogliamo stimare una probabilità, ci serve una qualche funzione che rimanga tra 0 e 1
- Introduciamo il sigmoide:

$$\sigma(a) \triangleq \frac{1}{1 + e^{-a}} \in (0, 1)$$

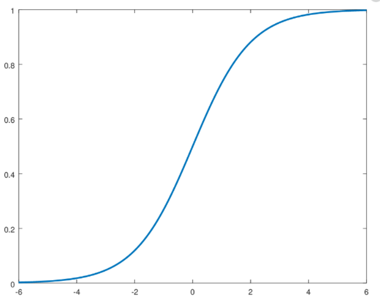


Sigmoide

$\Rightarrow$ Invece di stimare t, proviamo a stimare $P(t = 1|\mathbf{x})$, cioè la probabilità di ottenere un successo (cioè, $t = 1$), conoscendo i dati $\mathbf{x}$

- Dato che vogliamo stimare una probabilità, ci serve una qualche funzione che rimanga tra 0 e 1
- Introduciamo il sigmoide:

$$\sigma(a) \triangleq \frac{1}{1 + e^{-a}} \in (0, 1)$$



$\Rightarrow$ funzione *Logistica*

Odds

Con il termine *odds* si intende il rapporto tra la probabilità p di un evento e la probabilità $1 - p$ che tale evento non accada:

$$\text{odds}(p) \triangleq \frac{p}{1-p}$$

Odds

Con il termine *odds* si intende il rapporto tra la probabilità p di un evento e la probabilità $1 - p$ che tale evento non accada:

$$\text{odds}(p) \triangleq \frac{p}{1-p}$$

Esempio: Framingham Heart Study (1948)

Odds

Con il termine *odds* si intende il rapporto tra la probabilità p di un evento e la probabilità $1 - p$ che tale evento non accada:

$$\text{odds}(p) \triangleq \frac{p}{1-p}$$

Esempio: Framingham Heart Study (1948)

- Nella cittadina di Framingham (Massachusetts) furono esaminati nell'arco di 12 anni 656 pazienti nella fascia d'età 50-60 anni

Odds

Con il termine *odds* si intende il rapporto tra la probabilità p di un evento e la probabilità $1 - p$ che tale evento non accada:

$$\text{odds}(p) \triangleq \frac{p}{1-p}$$

Esempio: Framingham Heart Study (1948)

- Nella cittadina di Framingham (Massachusetts) furono esaminati nell'arco di 12 anni 656 pazienti nella fascia d'età 50-60 anni
- Questi 656 pazienti conducevano tutti uno stile di vita non particolarmente sano, soggetto a una dieta non equilibrata e al consumo di troppo alcool e/o tabacco

Odds

Con il termine *odds* si intende il rapporto tra la probabilità p di un evento e la probabilità $1 - p$ che tale evento non accada:

$$\text{odds}(p) \triangleq \frac{p}{1-p}$$

Esempio: Framingham Heart Study (1948)

- Nella cittadina di Framingham (Massachusetts) furono esaminati nell'arco di 12 anni 656 pazienti nella fascia d'età 50-60 anni
- Questi 656 pazienti conducevano tutti uno stile di vita non particolarmente sano, soggetto a una dieta non equilibrata e al consumo di troppo alcool e/o tabacco
- 130 di loro svilupparono una *coronaropatia*

Odds

Con il termine *odds* si intende il rapporto tra la probabilità p di un evento e la probabilità $1 - p$ che tale evento non accada:

$$\text{odds}(p) \triangleq \frac{p}{1-p}$$

Esempio: Framingham Heart Study (1948)

- Nella cittadina di Framingham (Massachusetts) furono esaminati nell'arco di 12 anni 656 pazienti nella fascia d'età 50-60 anni
- Questi 656 pazienti conducevano tutti uno stile di vita non particolarmente sano, soggetto a una dieta non equilibrata e al consumo di troppo alcool e/o tabacco
- 130 di loro svilupparono una *coronaropatia*
- La probabilità di sviluppare una coronaropatia è di $p = 130/656 = 0.2$

Odds

Con il termine *odds* si intende il rapporto tra la probabilità p di un evento e la probabilità $1 - p$ che tale evento non accada:

$$\text{odds}(p) \triangleq \frac{p}{1-p}$$

Esempio: Framingham Heart Study (1948)

- Nella cittadina di Framingham (Massachusetts) furono esaminati nell'arco di 12 anni 656 pazienti nella fascia d'età 50-60 anni
- Questi 656 pazienti conducevano tutti uno stile di vita non particolarmente sano, soggetto a una dieta non equilibrata e al consumo di troppo alcool e/o tabacco
- 130 di loro svilupparono una *coronaropatia*
- La probabilità di sviluppare una coronaropatia è di $p = 130/656 = 0.2$
- L'odds di sviluppare una coronaropatia è $\text{odds}(p) = \frac{0.2}{1-0.2} = 0.25$, cioè $1 : 4$

Regressione Logistica

L'assunzione chiave del metodo conosciuto come regressione logistica è che ci sia una dipendenza lineare tra i dati e il logaritmo degli odds:

$$\log\left(\frac{p}{1-p}\right) = \mathbf{w}^T\mathbf{x} = w_0 + w_1 x_1 + \ldots + w_K x_K$$

Regressione Logistica

L'assunzione chiave del metodo conosciuto come regressione logistica è che ci sia una dipendenza lineare tra i dati e il logaritmo degli odds:

$$\log\left(\frac{p}{1-p}\right) = \mathbf{w}^T\mathbf{x} = w_0 + w_1x_1 + \ldots + w_Kx_K$$

$$\frac{p}{1-p} = e^{\mathbf{w}^T\mathbf{x}}$$

Regressione Logistica

L'assunzione chiave del metodo conosciuto come regressione logistica è che ci sia una dipendenza lineare tra i dati e il logaritmo degli odds:

$$\log\left(\frac{p}{1-p}\right) = \mathbf{w}^T\mathbf{x} = w_0 + w_1x_1 + \ldots + w_Kx_K$$

$$\frac{p}{1-p} = e^{\mathbf{w}^T\mathbf{x}}$$

$$p = \frac{e^{\mathbf{w}^T\mathbf{x}}}{1+e^{\mathbf{w}^T\mathbf{x}}} = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}} = \sigma(\mathbf{w}^T\mathbf{x})$$

Regressione Logistica

L'assunzione chiave del metodo conosciuto come regressione logistica è che ci sia una dipendenza lineare tra i dati e il logaritmo degli odds:

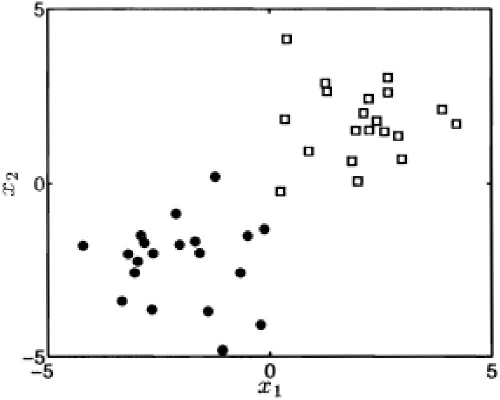
$$\log\left(\frac{p}{1-p}\right) = \mathbf{w}^T\mathbf{x} = w_0 + w_1x_1 + \ldots + w_Kx_K$$

$$\frac{p}{1-p} = e^{\mathbf{w}^T\mathbf{x}}$$

$$p = \frac{e^{\mathbf{w}^T\mathbf{x}}}{1+e^{\mathbf{w}^T\mathbf{x}}} = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}} = \sigma(\mathbf{w}^T\mathbf{x})$$

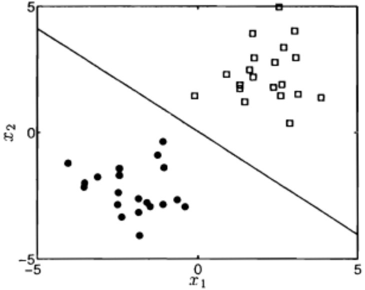
Per questo modello non c'è un modo di ottenere il regressore ottimo $\hat{\mathbf{w}}$ in formula chiusa. E'necessario utilizzare algoritmi di ottimizzazione (Gradient Method, Newton-Rhapson, ecc.)

Regressione Logistica



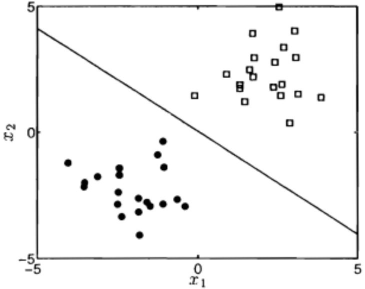
Assumiamo che i quadratini bianchi corrispondano all'evento con probabilità p

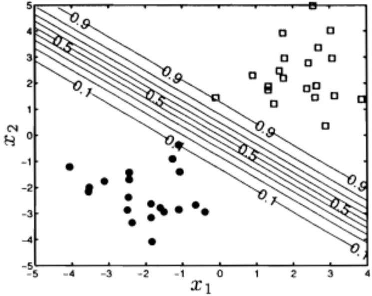
Regressione Logistica



$$P(t = 1|\mathbf{x}, \hat{\mathbf{w}}) = \sigma(\hat{\mathbf{w}}^T \mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2$$

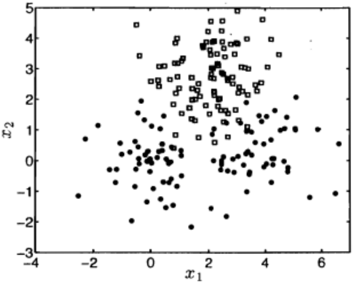
Regressione Logistica





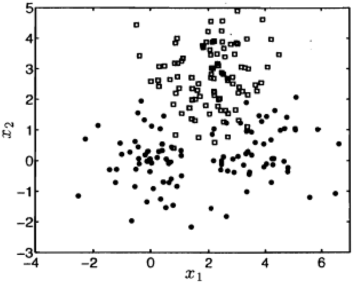
$$P(t = 1|\mathbf{x}, \hat{\mathbf{w}}) = \sigma(\hat{\mathbf{w}}^T \mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2$$

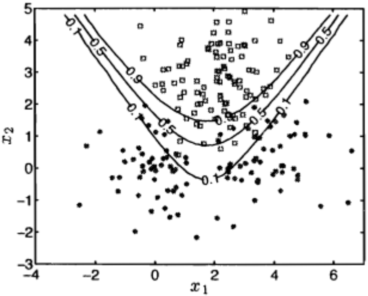
Regressione Logistica



$$P(t = 1|\mathbf{x}, \hat{\mathbf{w}}) = \sigma(\hat{\mathbf{w}}^T \mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2$$

Regressione Logistica





$$P(t = 1|\mathbf{x}, \hat{\mathbf{w}}) = \sigma(\hat{\mathbf{w}}^T \mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2$$