

# 16. Regressione logistica

Corso di Python per il Calcolo Scientifico

# Outline

- La regressione logistica
- La funzione sigmoidale
- La funzione di costo
- La regressione logistica in Scikit Learn

# La regressione logistica

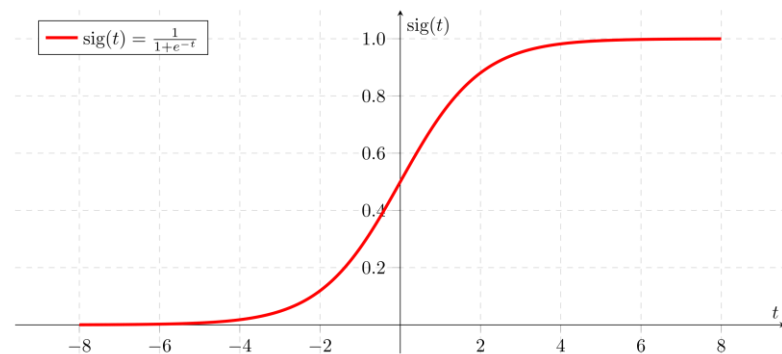
- Nonostante il suo nome, la **regressione logistica** è un classificatore
- Un modello di questo tipo **predice la probabilità che un dato appartenga ad una certa classe**
- Ad esempio, un modello di regressione logistica può predire che:
$$p(mail|unknown) = 0.05$$
- Ciò significa che la probabilità  $p(mail|unknown)$  che una mail ricevuta da un indirizzo sconosciuto sia di spam è pari a 0.05, ovvero che 5 mail su 100 ricevute da indirizzi sconosciuti saranno (mediamente) di spam

# La funzione sigmoidale

- La **funzione sigmoidale** viene usata da un regressore logistico (ma non solo) per assicurarsi che l'output di ciò che la precede ricada all'interno del range  $[0, 1]$
- La formulazione è la seguente:

$$y = \frac{1}{1 + e^{-t}}$$

- dove  $t = b + w_1x_1 + \dots + w_nx_n$  è un generico modello lineare



# La funzione di costo

- La funzione di costo usata nella regressione logistica è differente da quella utilizzata nella regressione lineare
- Viene chiamata **log loss**, ed è espressa come:

$$LogLoss = \sum_{(x,y) \in D} -y \log(y') - (1 - y) \log(1 - y')$$

- dove  $(x, y)$  sono le coppie feature/label relative al dataset  $D$ ,  $y$  è la label vera per un dato insieme di feature, mentre  $y'$  è il valore predetto

# La regressione logistica in Scikit Learn

- La regressione logistica è implementata in Scikit Learn mediante oggetti di classe **LogisticRegression()**
- Questi oggetti sono degli stimatori, per cui è possibile usarli in modo simile a come sono usati gli oggetti di classe **LinearRegressor()**

```
clf = LogisticRegression()  
clf.fit(X_train, y_train)  
y_pred = clf.predict(X_test)
```

# Domande?

42