

Metodi di ML: Obbiettivi e Paradigmi di Classificazione

Web Minig & Retrieval, a.a. 2015-16

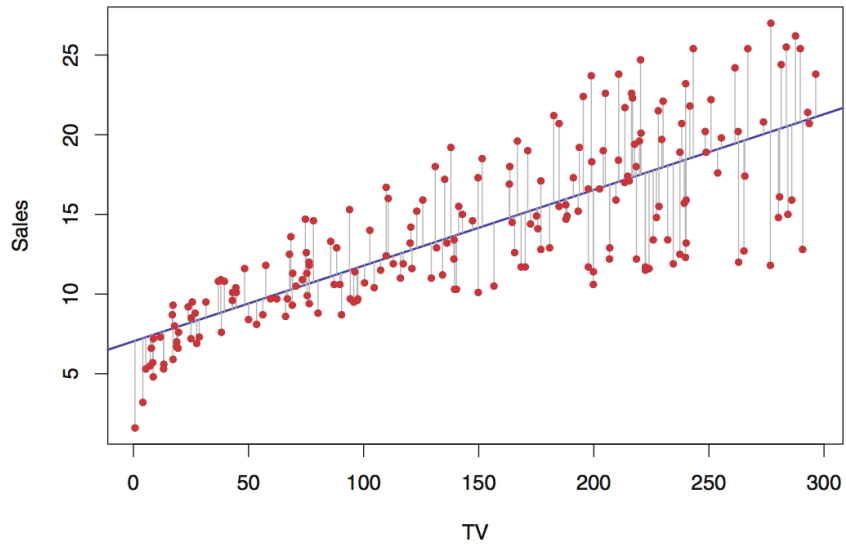
Roberto Basili

Sommario

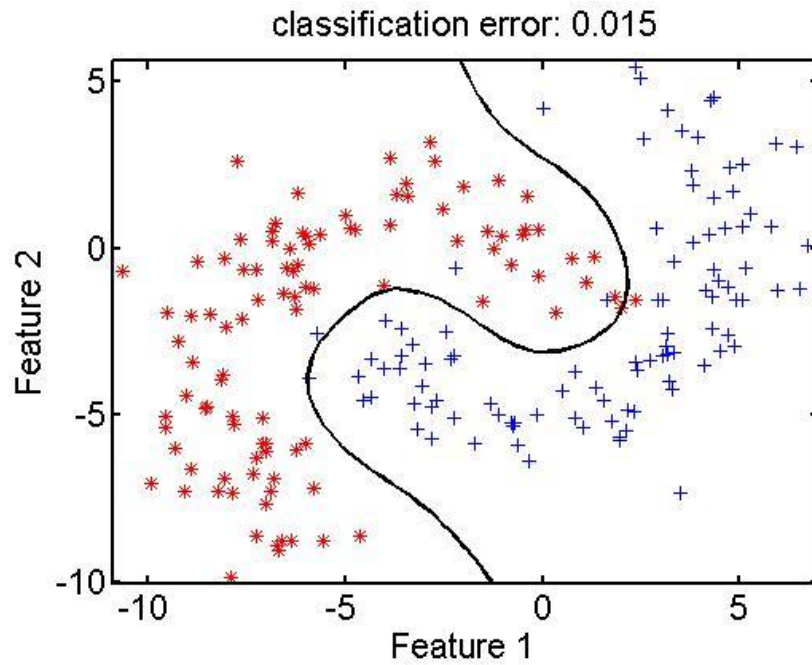
- Definizione dei problem target per il ML
- Paradigmi geometrici
- Paradigmi Probabilistici
 - Paradigmi Generativi

Machine Learning: the core problems

Regression



Classification



Machine Learning: the core problems

Regression

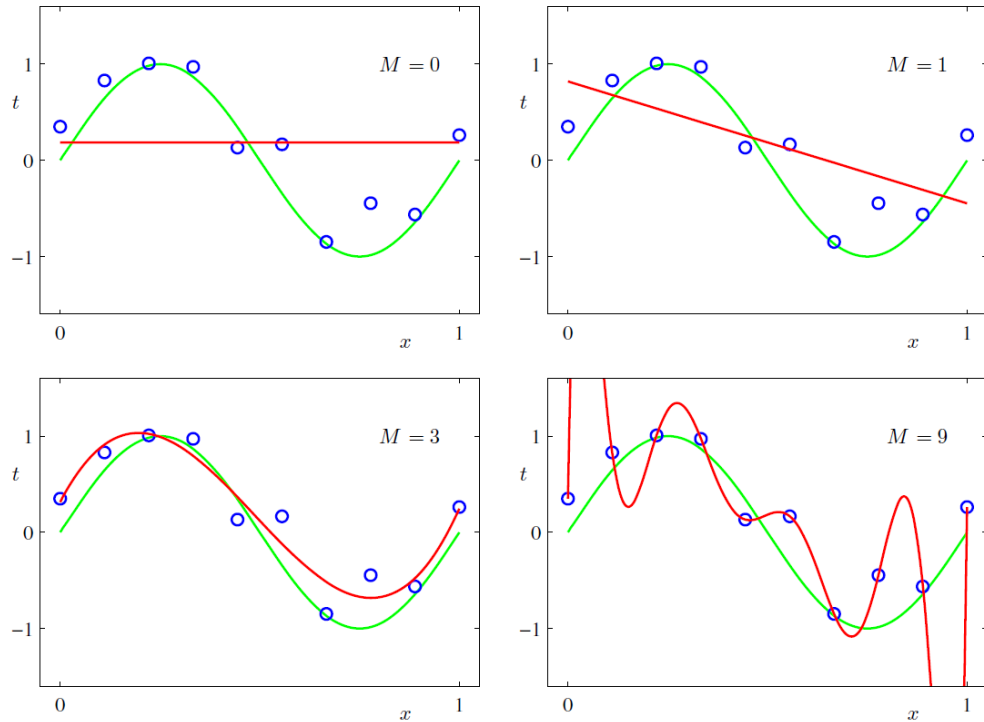
- Dati un insieme di esempi di una funzione $f(\cdot)$
- x_1, \dots, x_k con $y_i=f(x_i)$ nota per ogni i
- determinare un legge $h(\cdot)$ tale che:
 - $h(x_i) = y_i = f(x_i) \quad \forall i$
 - $h(x) \approx f(x)$ *altrove*

Classification

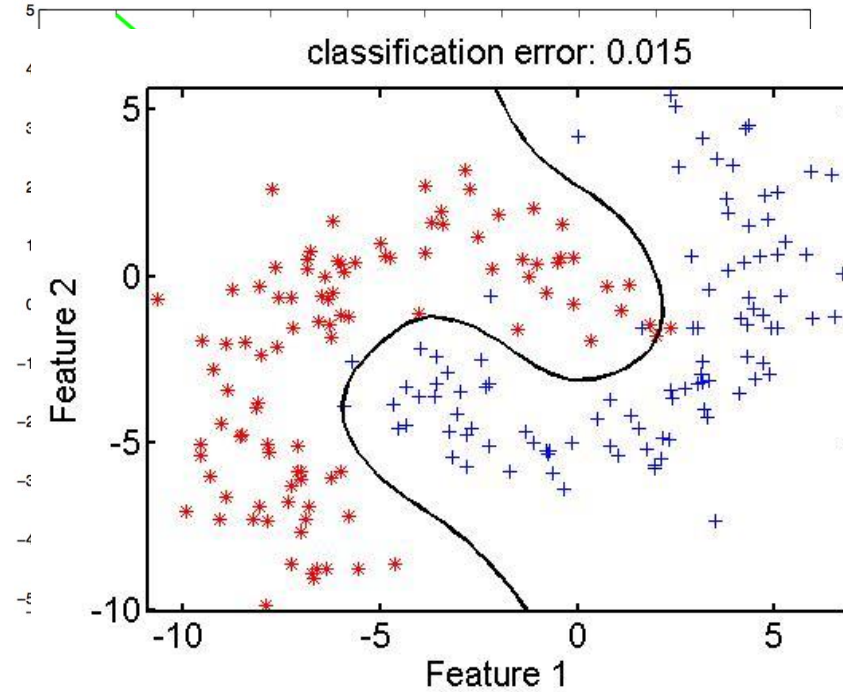
- Date n classi C_1, \dots, C_n e un certo numero di istanze x_1, \dots, x_k la cui classificazione y_1, \dots, y_k sia nota
- Determinare la legge $h(\cdot)$ tale che
 - $h(x_i) = y_i \quad \forall i=1, \dots, k$
 - $h(x)$ costituisca la classe corretta per ogni altra istanza x
-

Machine Learning: la scelta delle funzioni

Regression



Classification



Paradigmi di selezione del modello

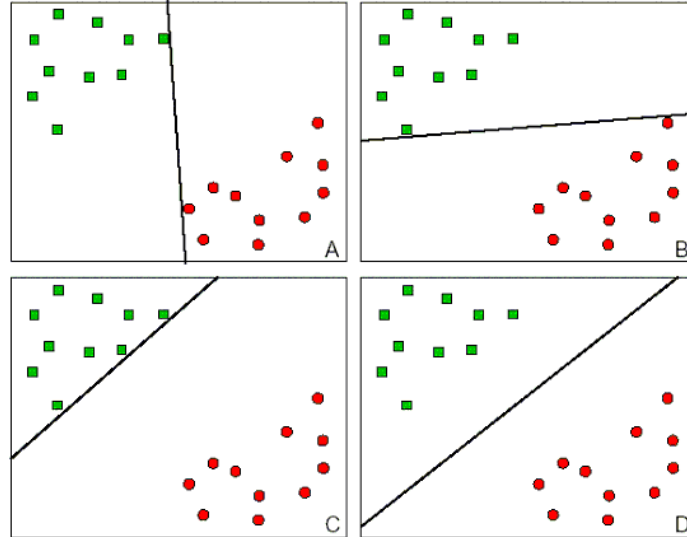
- La selezione del modello dipende dalla scelta
 - **(Model Family Selection)** Di una classe di funzioni (ad es. polinomi di grado n)
 - **(Model parametrization)**. Dalla selezione dei parametri usati per specificare la funzione ottima
 - Definizione del criterio di ottimalità (ad esempio *copertura* vs. *accuracy*)
 - Ricerca dei parametri ottimi
 - Analiticamente
 - Sfruttando il training set

Selezione della famiglia di funzioni

- Approcci discriminativi

- Lineari

- $h(\mathbf{x}) = \text{sign}(\mathbf{W} \cdot \mathbf{x} + \mathbf{b})$

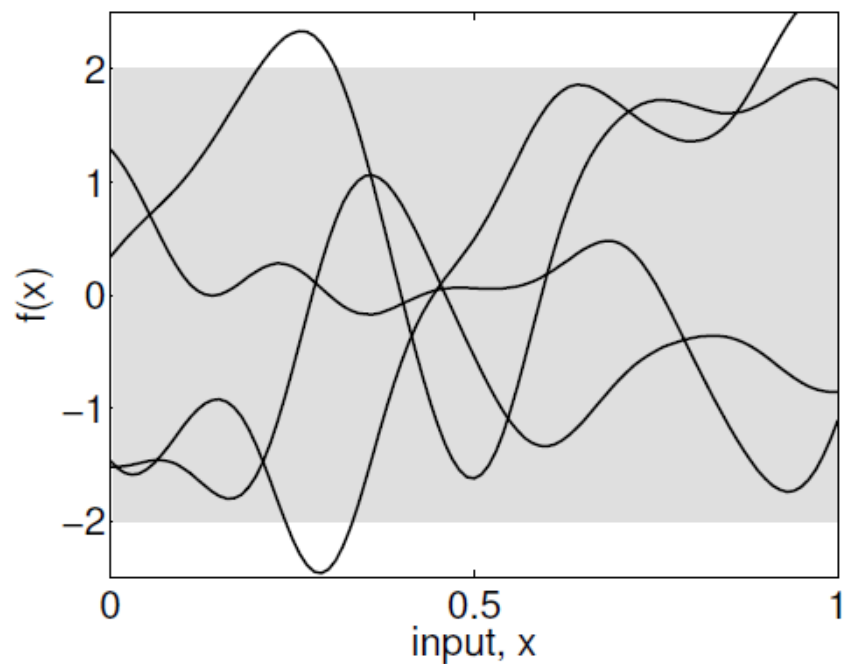


- Approcci probabilistici

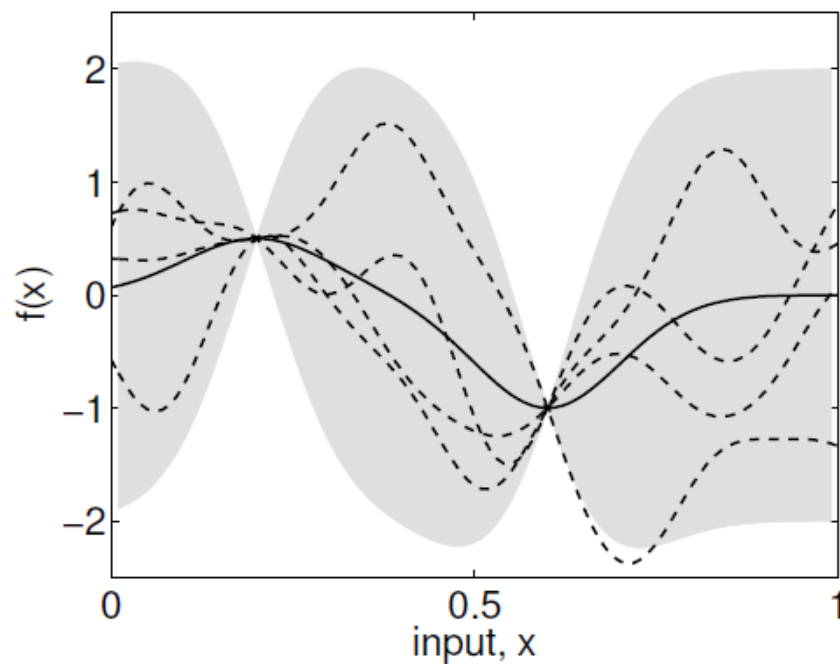
- Stima delle probabilità $p(\mathcal{C}_k|\mathbf{x})$ attraverso un training set
- Modello generativo ed uso della inversione Bayesiana

$$p(\mathcal{C}_k|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_k)p(\mathcal{C}_k)}{p(\mathbf{x})}.$$

Modelli Bayesian & Grafici

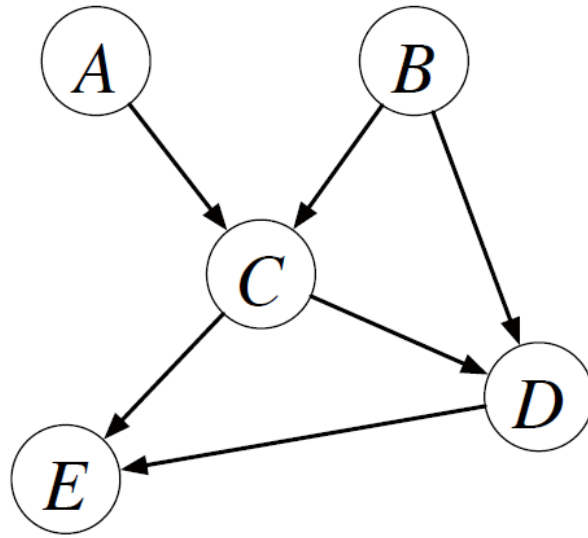


(a), prior



(b), posterior

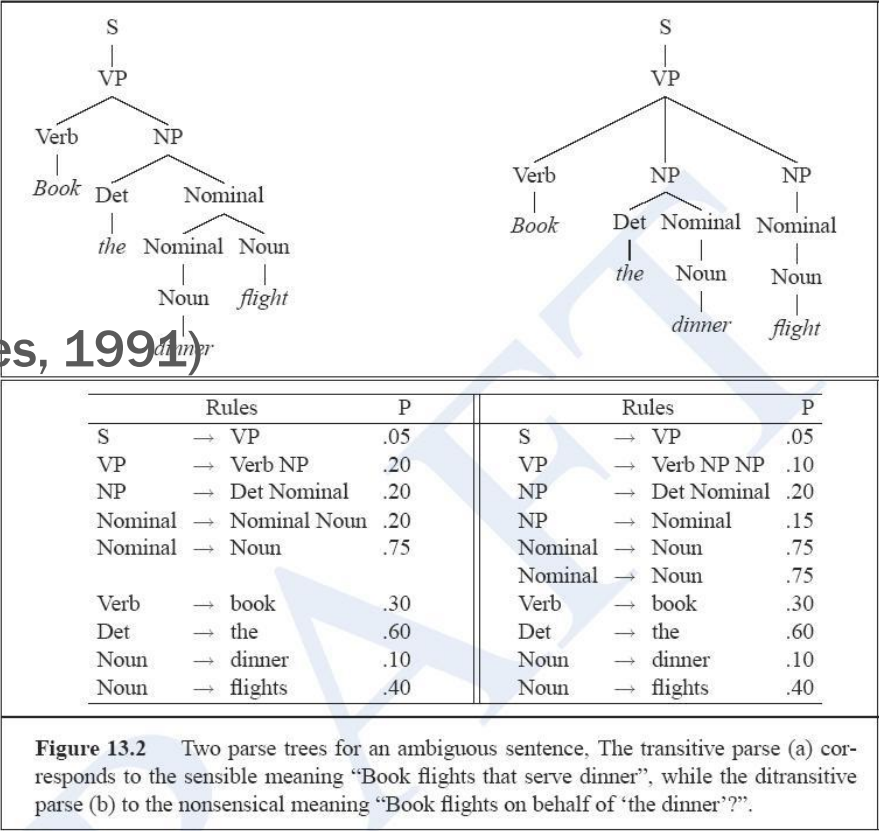
Modelli Grafici



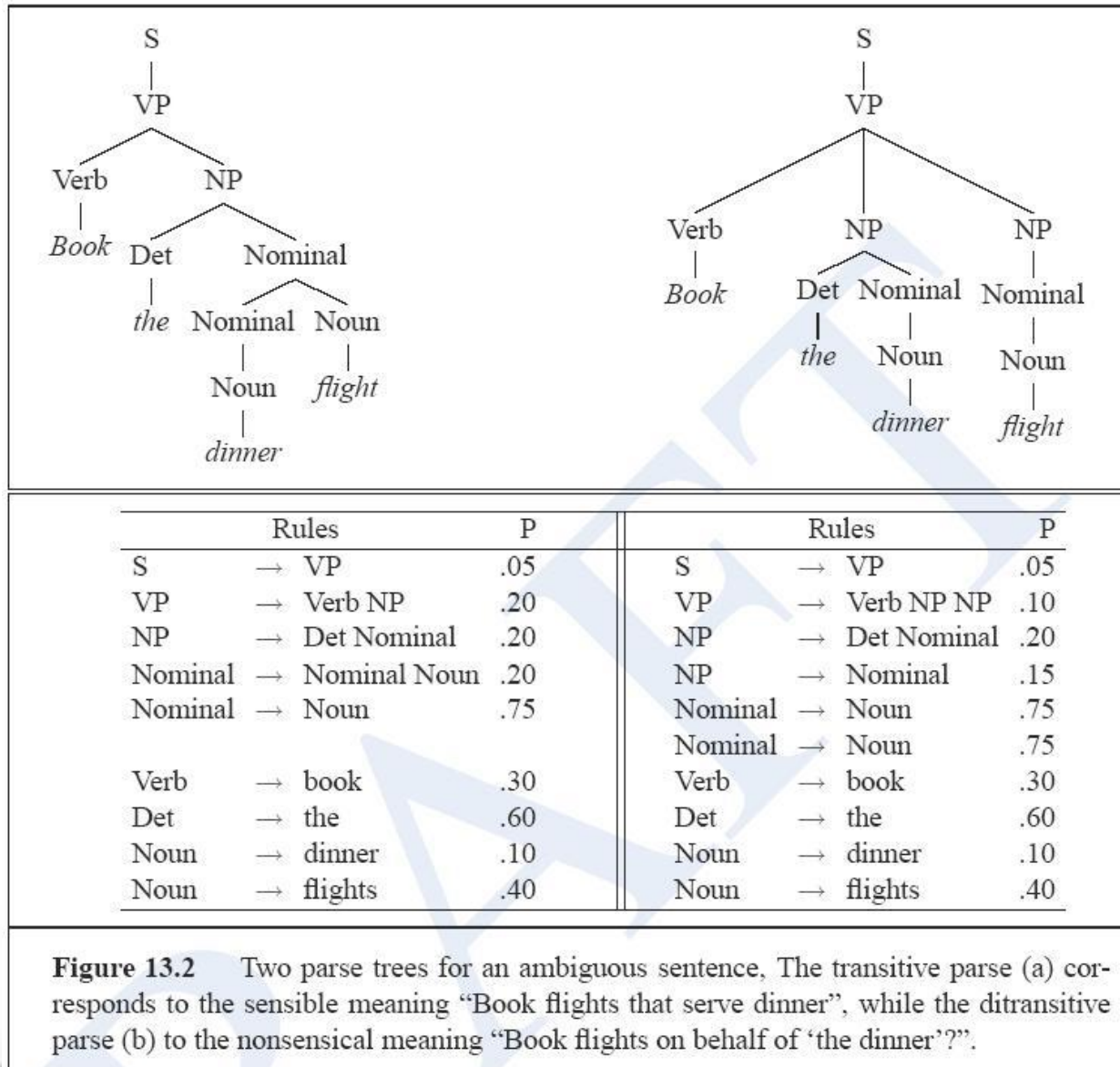
$$p(A, B, C, D, E) = p(A)p(B)p(C|A, B)p(D|B, C)p(E|C, D)$$

Weighted Grammars, tra Sintassi & Statistica

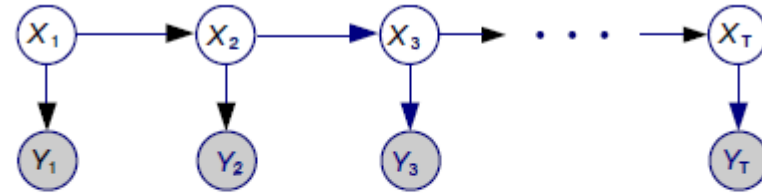
- POS tagging (Curch, 1989)
- Probabilistic Context-Free Grammars (Pereira & Schabes, 1991)
- Data Oriented Parsing (Scha, 1990)
- Stochastic Grammars (Abney, 1993)
- Lessicalizzati Modelli (C. Manning, 1995)



Weighted Grammars, tra Sintassi & Statistica



Hidden Markov Models



- Stati = Categorie/Concetti/Proprietà
- Osservazioni: simboli di un certo linguaggio
- Emissioni vs. Transizioni
- Applicazioni:
 - Speech Recognition (simboli:fonemi, stati:segmentazione)
 - POS tagging (simboli: parole, stati: categorie gramaticali)

$$p(X_1, \dots, X_T, Y_1, \dots, Y_T) = p(X_1) p(Y_1 | X_1) \prod_{t=2}^T [p(X_t | X_{t-1}) p(Y_t | X_t)]$$