

## Capitolo 8

### 1 Apprendimento Probabilistico

#### 1.1 Modelli di Apprendimento Probabilistico

- **Modello MAP (Maximum A Posteriori):**

- Combina la probabilità a priori  $P(h)$  e la verosimiglianza  $P(D | h)$  per trovare l'ipotesi  $h$  che massimizza la probabilità posteriore:

$$h_{MAP} = \operatorname{argmax}_h P(h | D) = \operatorname{argmax}_h P(D | h) \cdot P(h).$$

- Vantaggi: incorpora conoscenze pregresse tramite  $P(h)$ .

- **Modello ML (Maximum Likelihood):**

- Trova l'ipotesi  $h$  che massimizza la verosimiglianza dei dati:

$$h_{ML} = \operatorname{argmax}_h P(D | h).$$

- Non considera informazioni a priori ( $P(h)$ ).

#### 1.2 Apprendere Probabilità

- **Pseudo-conteggi:**

- Metodo per stimare probabilità da piccoli campioni o dati incompleti.
- Usa valori iniziali "fittizi" per evitare probabilità nulle:

$$P(X = x) = \frac{\text{conteggio}(X = x) + \alpha}{N + \alpha \cdot k},$$

dove  $\alpha$  è un parametro che controlla l'influenza degli pseudo-conteggi.

- **Opinione degli Esperti:**

- Combina conoscenze empiriche e probabilistiche con giudizi esperti.

#### 1.3 Classificatori Probabilistici

- **Naive Bayes:**

- Assume che le feature  $X_1, X_2, \dots, X_n$  siano condizionatamente indipendenti dato la classe  $Y$ .
- Classifica un'istanza calcolando:

$$P(Y | X_1, X_2, \dots, X_n) \propto P(Y) \cdot \prod_{i=1}^n P(X_i | Y).$$

- **Apprendimento di Classificatori Bayesiani:**

- Si estende Naive Bayes incorporando dipendenze tra le feature.

## 1.4 Apprendimento MAP di Alberi di Decisione

- Combina alberi di decisione con probabilità a priori.
- Minimizza la lunghezza delle descrizioni dei dati osservati e del modello (principio MDL - Minimum Description Length).

## 2 Apprendimento Non Supervisionato

### 2.1 k-Means

- **Definizione:**
  - Algoritmo iterativo che partiziona i dati in  $k$  cluster, minimizzando la distanza intra-cluster.
  - Passaggi:
    - \* Assegna ogni punto al centroide più vicino.
    - \* Calcola nuovi centroidi come media dei punti assegnati.
- **Applicazioni:**
  - Segmentazione dei dati, compressione, clustering.

### 2.2 Expectation-Maximization (EM) e Soft Clustering

- **EM:**
  - Algoritmo per massimizzare la verosimiglianza in presenza di dati incompleti o incerti.
  - Due fasi:
    - \* **Expectation (E-step):** calcola le probabilità attese per ogni assegnazione.
    - \* **Maximization (M-step):** aggiorna i parametri massimizzando la probabilità.
- **Soft Clustering:**
  - Ogni punto appartiene a più cluster con probabilità associate, anziché assegnazioni rigide.

## 3 Apprendimento di Belief Network

### 3.1 Imparare le Probabilità

- Utilizza dati osservati per stimare le probabilità condizionali tra variabili.

### 3.2 Variabili Latenti e Dati Mancanti

- **Variabili Latenti:** Variabili non osservate direttamente, ma che influenzano il modello.
- **Dati Mancanti:** EM è spesso utilizzato per stimare i valori mancanti.

### 3.3 Apprendimento della Struttura

- **Metodo Search & Score:** Esplora lo spazio delle possibili reti usando un criterio di punteggio (es., MDL, verosimiglianza).
- **Caso Generale:** Combina vincoli esperti e dati osservati per ottimizzare la struttura.

## 4 Apprendimento Bayesiano

- Combina in modo formale conoscenze pregresse e dati osservati.
- Aggiorna costantemente la distribuzione delle ipotesi usando il Teorema di Bayes:

$$P(h \mid D) = \frac{P(D \mid h) \cdot P(h)}{P(D)}.$$

## Conclusioni

Questi appunti forniscono una panoramica strutturata dei metodi di apprendimento probabilistico e non supervisionato. Si enfatizzano tecniche come Naive Bayes, EM, e il clustering  $k$ -means, oltre alla costruzione e all'apprendimento delle Belief Network, evidenziando l'integrazione tra teoria e applicazioni pratiche.