**Sintesi Completa ed Esaustiva degli Appunti: Apprendimento Probabilistico e Non Supervisionato**

**1. Apprendimento Probabilistico**

**1.1 Modelli di Apprendimento Probabilistico**

1. **Modello MAP (Maximum A Posteriori):**
   * Combina la probabilità a priori P(h)P(h) e la verosimiglianza P(D∣h)P(D | h) per trovare l'ipotesi hh che massimizza la probabilità posteriore: hMAP=argmaxhP(h∣D)=argmaxhP(D∣h)⋅P(h).h\_{MAP} = \text{argmax}\_h P(h | D) = \text{argmax}\_h P(D | h) \cdot P(h).
   * Vantaggi: incorpora conoscenze pregresse tramite P(h)P(h).
2. **Modello ML (Maximum Likelihood):**
   * Trova l'ipotesi hh che massimizza la verosimiglianza dei dati: hML=argmaxhP(D∣h).h\_{ML} = \text{argmax}\_h P(D | h).
   * Non considera informazioni a priori (P(h)P(h)).

**1.2 Apprendere Probabilità**

1. **Pseudo-conteggi:**
   * Metodo per stimare probabilità da piccoli campioni o dati incompleti.
   * Usa valori iniziali "fittizi" per evitare probabilità nulle: P(X=x)=conteggio(X=x)+αN+α⋅k,P(X=x) = \frac{\text{conteggio}(X=x) + \alpha}{N + \alpha \cdot k}, dove α\alpha è un parametro che controlla l'influenza degli pseudo-conteggi.
2. **Opinione degli Esperti:**
   * Combina conoscenze empiriche e probabilistiche con giudizi esperti.

**1.3 Classificatori Probabilistici**

1. **Naive Bayes:**
   * Assume che le feature X1,X2,…,XnX\_1, X\_2, \ldots, X\_n siano condizionatamente indipendenti dato la classe YY.
   * Classifica un'istanza calcolando: P(Y∣X1,X2,…,Xn)∝P(Y)⋅∏i=1nP(Xi∣Y).P(Y | X\_1, X\_2, \ldots, X\_n) \propto P(Y) \cdot \prod\_{i=1}^n P(X\_i | Y).
2. **Apprendimento di Classificatori Bayesiani:**
   * Si estende Naive Bayes incorporando dipendenze tra le feature.

**1.4 Apprendimento MAP di Alberi di Decisione**

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

**2. Apprendimento Non Supervisionato**

**2.1 k-Means**

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

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

1. **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.
2. **Soft Clustering:**
   * Ogni punto appartiene a più cluster con probabilità associate, anziché assegnazioni rigide.

**3. Apprendimento di Belief Network**

**3.1 Imparare le Probabilità**

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

**3.2 Variabili Latenti e Dati Mancanti**

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

**3.3 Apprendimento della Struttura**

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

**4. Apprendimento Bayesiano**

1. Combina in modo formale conoscenze pregresse e dati osservati.
2. Aggiorna costantemente la distribuzione delle ipotesi usando il Teorema di Bayes: P(h∣D)=P(D∣h)⋅P(h)P(D).P(h | D) = \frac{P(D | 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 kk-means, oltre alla costruzione e all'apprendimento delle Belief Network, evidenziando l'integrazione tra teoria e applicazioni pratiche.